

2012

Stability (over time) of regularized modified CS (noisy) for recursive causal sparse reconstruction

Fardad Raisali
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Electrical and Electronics Commons](#)

Recommended Citation

Raisali, Fardad, "Stability (over time) of regularized modified CS (noisy) for recursive causal sparse reconstruction" (2012). *Graduate Theses and Dissertations*. 12973.
<https://lib.dr.iastate.edu/etd/12973>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

**Stability (over time) of regularized modified CS (noisy) for recursive causal
sparse reconstruction**

by

Fardad Raisali

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Major: Electrical Engineering

Program of Study Committee:
Namrata Vaswani, Major Professor
Zhendao Wang
Mani Mina

Iowa State University

Ames, Iowa

2012

Copyright © Fardad Raisali, 2012. All rights reserved.

DEDICATION

I'd like to dedicate this work to my grandfather who will be always in my heart.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	iv
ACKNOWLEDGEMENTS	v
ABSTRACT	vi
CHAPTER 1. Introduction	1
1.1 Motivation	1
1.2 Notation and problem definition	1
1.3 Past works and our contribution	2
1.4 Thesis Organization	5
CHAPTER 2. Sparse Signal Recovery from Noisy Measurements with Partial Knowledge of Signal Support and Value	7
2.1 Regularized Modified CSN	7
2.2 Regularized Modified-BPDN	8
CHAPTER 3. Signal Models for Sparse and Compressible Sequential Signals	11
3.1 Signal Model I	11
3.2 Signal Model II	13
3.2.1 Discussion of Signal Model	13
3.3 Comparison of signal model I and II	15
CHAPTER 4. Algorithm for Sequential Reg-mod-CSN and Stability Results	17
4.1 Algorithm for Sequential Regularized Modified CSN over time	17

4.2	Stability Results with Signal Model I	17
4.2.1	Discussion of Theorem	19
4.2.2	Simulation Result	20
4.3	Stability Results with Signal Model II	21
4.3.1	Discussion of Theorem	23
4.3.2	Simulation Result	23
CHAPTER 5. Summary and Future Work		27
APPENDIX A. APPENDIX		28
A.1	Proof of Theorem 1	28
A.2	Proof of Lemma 2	31
A.3	Proof of Theorem 2	36
A.4	Proof of Theorem 4	40
BIBLIOGRAPHY		42
BIBLIOGRAPHY		42

LIST OF FIGURES

3.1	(a) Sorted MRI signal coefficient values in Wavelet domain (b) Sorted proposed signal model coefficient values in Wavelet domain	15
4.1	Normalized MSE (NMSE), number of extras and number of misses over time for CS, modified CS, modified BPDN, and regularized modified CSN. In part (b) and (c), NMSE for CS was more than 20%.(plotted only in (a)	21
4.2	Normalized MSE (NMSE), number of extras and number of misses over time for CS, modified CS, modified BPDN, and regularized modified CSN generated signal by signal model. In part (b) and (c), NMSE for CS was more than 20%.(plotted only in (a)	26

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

I would like to express my deepest gratitude to my advisor, Prof. Namrata Vaswani, for her excellent guidance, caring, patience. This work was not possible without her help and support.

I would like to thank Prof.Mina, Prof.Manimaran, Prof.Wang for guiding and supporting me within the time I was studying at Iowa State University.

Many thanks to my lab colleges, Samarjit, Wei, Taoran and Chenlu. We had a great time of being together.

I would also like to thank my parents and sister that they were always supporting me and encouraging me with their best wishes.

ABSTRACT

Accurate signal recovery from under-determined system of equations is a topic of considerable interest. Compressive sensing (CS) gives an approach to find a solution to this system when the unknown signal is sparse. Regularized modified CS (noisy) propose an approach to find the solution to the under-determined system of equations when we are provided with 1- Partial part of signal support denoted by T and 2- A prior estimate of signal value on this support denoted by μ_T . In many applications, e.g sequential MRI reconstruction, the sparse signal support and its nonzero signal values change slowly over time. Inspired by this fact, we propose an algorithm utilizing reg-mod-CSN for sequential signal reconstruction such that the prior estimate of T and μ_T is generated from the previous time instant.

Our major focus in this work is to study the "stability" of the proposed algorithm for recursive reconstruction of sparse signal sequences from noisy measurements. By "stability" we mean that the number of misses from the current support estimate; the number of extras in it; and the ℓ_2 norm of the reconstruction error remain bounded by a *time-invariant* value at all times. For achieving this goal, we need a signal model that can represent the sequential signals in real applications. It should satisfy three constraint; 1- The distribution of the signal entries should follow the same distribution as real sequential signals; 2- It follows the same evolutionary pattern as the real sequential signals over time and 3-The signal support changes dynamically over time. In the two proposed signal model, we tried to satisfy these three constraints. Using these signal models, we analyzed the performance of the proposed algorithm and found the condition such that the system remain stable. These conditions are weaker in compare with older methods like CS and mod-CS. At the end, we show empirically that reg-mod-CS achieves a lower reconstruction error in compare with mod-CS and CS.

CHAPTER 1. Introduction

1.1 Motivation

Recovering sparse signal from under-determined system of equations is a topic of interest in many areas. In medical applications (e.g. MRI imaging, CT imaging) fast signal recovery is quiet important. The time that the system spends on data acquisition is inappropriate in a sense that expose the patient with radiation for a longer time or make a poor reconstruction because of motion artifact. There are cases that one need to scan a particular part of body over a time sequence, e.g. sequential MRI imaging. In these cases the object may have small variations over time. This limited variation spark the idea of using past information to reconstruct the current time signal. If this initial information help to reconstruct the signal with lower number of measurements, that can lead to a gain in scanning time.

1.2 Notation and problem definition

The set operations \cup , \cap , \setminus have their usual meanings. \emptyset denotes the empty set. We use T^c to denote the complement of a set T w.r.t. $[1, m] := [1, 2, \dots m]$, i.e. $T^c := [1, m] \setminus T$. $|T|$ denotes the cardinality of T . For a vector, v , and a set, T , v_T denotes the $|T|$ length sub-vector containing the elements of v corresponding to the indices in the set T . $\|v\|_k$ denotes the ℓ_k norm of a vector v . *If just $\|v\|$ is used, it refers to $\|v\|_2$.* Similarly, for a matrix M , $\|M\|_k$ denotes its induced k -norm, while just $\|M\|$ refers to $\|M\|_2$. M' denotes the transpose of M . For a fat matrix A , A_T denotes the sub-matrix obtained by extracting the columns of A corresponding to the indices in T .

We obtain an n -length measurement vector y_t by

$$y_t = Ax_t + w_t$$

A is an $m \times n$ ($m \leq n$) matrix that we call the measurement matrix. x_t is an n length sparse vector with support N_t , y_t is the n length observation vector and w_t is the m length noise observation vector with $\|w_t\| \leq \rho$. We assume partial knowledge of support and denote it by T_t . Also we assume partial knowledge of the signal estimate on T_t , and denote it by $(\mu_t)_{T_t}$. The signal estimate is assumed to be zero along T_t^c .

Our goal is to recursively estimate x_t using y_1, \dots, y_t . By recursively, we mean, use only y_t and the estimate from $t-1$, \hat{x}_{t-1} to compute the estimate at time t . Recursive recovery ensures both computational and storage complexity remains the same as that of simple CS (CS done for each time instant separately).

The S-restricted isometry constant δ_S , for a matrix, A , proposed in [1], is defined as the smallest positive number satisfying

$$(1 - \delta_S)\|c\|^2 \leq \|A_T c\| \leq (1 + \delta_S)\|c\|^2$$

for all subsets of T with cardinality $|T| \leq S$ and all real vectors c of length $|T|$. The S, S' restricted orthogonality constant, $\theta_{S, S'}$, proposed in [1], is defined as the smallest real number satisfying

$$\langle A_{T_1} c_1, A_{T_2} c_2 \rangle \leq \theta_{S, S'} \|c_1\| \|c_2\|$$

for all disjoint sets T_1, T_2 with $|T_1| \leq S$, $|T_2| \leq S'$ and $S + S' \leq m$, and for all vectors c_1, c_2 of length $|T_1|, |T_2|$ respectively.

Definition 1 ($T_t, \Delta_t, \Delta_{e,t}$). We use T_t to denote the support estimate at time t from the previous time $t-1$. We use $\Delta_t := N_t \setminus T_t$ to denote the unknown part of the support estimate and $\Delta_{e,t} := T_t \setminus N_t$ to denote the “erroneous” part of T_t .

Definition 2 ($\tilde{T}_t, \tilde{\Delta}_t, \tilde{\Delta}_{e,t}$). We use \tilde{T}_t to denote the final support estimate at current time t . We use $\tilde{\Delta}_t := N_t \setminus \tilde{T}_t$ and $\tilde{\Delta}_{e,t} := \tilde{T}_t \setminus N_t$.

1.3 Past works and our contribution

The problem of recovering unknown signal from under-determined system of equations has attained a lot of attention in recent years. Compressive sensing (CS) is a novel approach which

direct this problem for sparse or compressible signals([1], [2], [3], [4]). We obtain an n -length measurement vector y by

$$y = Ax + w$$

A is an $m \times n$ ($m \leq n$) matrix that we call the measurement matrix. x is an n length sparse vector, y is the m length observation vector and w is the m length noise vector such that $\|w\| \leq \rho$.

One approach for sparse signal reconstruction is to consider all possible signals that satisfy certain level of sparsity and searching the true signal among all of these candidates. In other word, we are looking for the sparsest solution such that satisfy the data constraint

$$\min \|\beta\|_0 \quad s.t \quad \|y - A\beta\| \leq \rho$$

This problem gives the true solution if $\delta_{|N|} < 1$ which N represents the support set of signal x and $|N|$ denotes the cardinality of this set. Actually this problem is NP-hard and interactive. In [5] for the noiseless case, it was shown that under certain conditions, this problem can be solved via convex relaxation. Compressive sensing attempts to reconstruct sparse signal x , by solving

$$\min \|\beta\|_1 \quad s.t \quad \|y - A\beta\| \leq \rho$$

the succession of the optimization problem depends on matrix A and sparsity level of signal x . More precisely, in the noiseless case, this problem gives the exact solution if $\delta_T < \frac{\sqrt{2}-1}{2}$. For the noisy case, an error bound was proposed in [6], which gives the reconstruction error proportional to ρ .

There are many algorithms proposed for solving these two optimization problems ([7], [8], [9], [10], [4], [11]). Classical CS assume that we are not provided with any prior information about the signal value and signal support. But in some cases, the observer has been provided with some information about the signal. There are many works that tried to employ this prior information ([12], [13], [14], [15]). Suppose that we are provided with partial part of the signal support $T \subset N$ where N is the signal support. Modified-CS(mod-CS) [16] tries to find a signal that is sparsest outside of T and satisfies the data constraint. It tries to reconstruct signal x ,

by solving

$$\min \|\beta_{T^c}\|_1 \quad s.t. \quad \|y - A\beta\| \leq \rho$$

Mod-CS shows that sparse signal, x , is recoverable under much weaker conditions in compare with CS. More precisely, for the noiseless case, it gives the exact solution if $\delta_{|T|+2|\Delta|} < \frac{\sqrt{2}-1}{2}$. In addition to the partial knowledge of support, it is possible that we have been provided with an initial guess of the signal value on this support, μ_T , such that $\|x_T - \mu_T\| \leq \gamma$ where γ is an scalar.

In this work [17], we first propose two convex optimization problem, reg-mod-CSN and reg-mod-BPDN, to use these extra information for signal recovery. Regularized modified CS(noisy)(reg-mod-CSN) is the noisy relaxation of regularized modified CS (reg-mod-cs) proposed in [16]. Reg-mod-CSN and reg-mod-BPDN try to find a signal that is sparsest outside of T ; is "close" enough to μ_T on T ; and satisfies the data constraint.

There are many examples in sequential signals that both the sparse signal's support and its nonzero signal values change slowly over time. This assumption has been empirically verified in earlier work [16] for medical image sequences. Using this characteristic, we propose an algorithm that utilize reg-mod-csn for sparse reconstruction over time. At each time instant, it gives an initial guess about the current time signal support and value by using the reconstructed signal form the previous time.

Other algorithms for recursive reconstruction include our older work on Least Squares CS-residual (LS-CS) and Kalman filtered CS-residual (KF-CS) [18–20]; modified-CS [16]; homotopy methods [21] (use past reconstructions to speed up current optimization but not to improve reconstruction error with fewer measurements); and [22] (a recent modification of KF-CS). Another recent work on CS for time-varying signals [23] proposed a series of causal but batch approaches that assume a *time-invariant* support.

Two other algorithms that are also designed for static CS with partial knowledge of support include [24] and [25]. The work of [24] proposed an approach similar to modified-CS but did not analyze it and also did not show real experiments either. The work of [25], which appeared in parallel with modified-CS, assumed a probabilistic prior on the support.

The proposed recursive algorithm estimate x_t using y_t and the estimate from $t - 1$, \hat{x}_{t-1} to

compute the estimate at time t . So the current reconstruction error depends on how well the previous time signal was estimated. In the other word, we like the reconstruction error to remain bounded by a time independent. Otherwise, the initial guess may mislead the reconstruction process in a way that the solution goes far from the true signal. In this work, we study the "stability" of Regularized modified CS(noisy) for recursive reconstruction of sparse signal sequences from noisy measurements. By "stability" we mean that the number of misses from the current support estimate; the number of extras in it; and the ℓ_2 norm of the reconstruction error remain bounded by a *time-invariant* value at all times. The concept is meaningful only if the support error bounds are small compared to the signal support size.

To the best of our knowledge, stability of recursive sparse reconstruction algorithms has not been studied in any other work except in older works [20, 26] for LS-CS and modified-CS respectively. The limitation of the result of [20] was that it assumed a signal model where support changes are only allowed every-so-often. But this assumption often does not hold in practice, e.g. for dynamic MRI sequences, support changes occur at every time. This limitation was removed in [26] where we used a signal model that allows support changes at every time t . In this work, first we use the same signal model to get the stability results.

The signal model in [26] generate a signal which it's energy at each time instant depends on the evolution of the signal over time. In this work, we propose another signal model that separate the distribution of signal vectors entries at each time instant from the evolution of the entries over time. Under this new signal model, we also find the conditions which the system remain stable. Our overall approach is also motivated by that of [26] for modified-CS. But there are significant differences since for reg-mod-CSN, the current reconstruction also depends on the previously reconstructed signal values (not just its support estimate), which makes its stability analysis more difficult.

1.4 Thesis Organization

In chapter 2, we introduce reg-mod-CSN and reg-mod-BPDN which are two optimization problems for recovering the unknown signal. In chapter 3, we propose two signal models. These models include the distribution of the signal vector values at each time instant and its

evolution over time. In chapter 4, an algorithm is introduced for applying reg-mod-CSN over a time sequence. Furthermore, the stability of the algorithm output is analyzed. At the end, conclusion is brought in chapter 5.

CHAPTER 2. Sparse Signal Recovery from Noisy Measurements with Partial Knowledge of Signal Support and Value

In this chapter we propose two optimization problem, reg-mod-CSN and reg-mod-BPDN, for signal recovery via noisy measurements with partial knowledge of signal support. The conditions where these convex problems have unique solution have been obtained. It also gives the l_2 reconstruction error for these two method which are mostly based on the same procedure in [27] and [5].

2.1 Regularized Modified CSN

In this section, we introduce regularized Modified CSN and derive the bound for its reconstruction error. We consider the case where there is one measurement vector, y , and a signal vector x .

$$y := Ax + w, \quad \text{where } \|w\| \leq \epsilon$$

Let N denote the support of x , i.e $N := \{i : |x_i| > \beta\}$ where $\beta \in \mathbb{R}$. Assume that we know partial part of support denoted by T . We define $\Delta = N \setminus T$. In addition to the measurements and partial knowledge of signal support, T , we know that signal x satisfies

$$\|x_T - \mu_T\|_2 \leq \gamma$$

where μ_T is the partial knowledge of the signal estimate on T . Regularized Modified CSN solves the following problem.

$$\min \|\beta_{T^c}\|_1 \text{ s.t } \|y - A\beta\|_2 \leq \epsilon \text{ and } \|\beta_T - \mu_T\|_2 \leq \gamma \quad (2.1)$$

The following theorem gives the sufficient conditions where reg-mod-csn have a unique solution and an reconstruction error bound for this solution

Theorem 1. Let $u := |T|$ and $k := |\Delta|$. Assume that $\delta_u < 1$, $1 - \delta_{2k} - \theta_{k,2k} > 0$ and $\|x_T - \mu_T\| \leq \gamma$. Then the solution \hat{x} to (2.1) obeys

$$\|x - \hat{x}\| \leq C_{u,k}\epsilon + D_{u,k}\gamma + E_{u,k}e_0(T, \Delta) \quad (2.2)$$

where

$$\begin{aligned} C_{u,k} &= 2 \frac{\sqrt{1 + \delta_u}}{1 - \delta_u} + 2 \frac{(2 + \frac{(\sqrt{2}+1)\theta_{u,k}}{1-\delta_u})\sqrt{1 + \delta_{2k}}}{1 - \delta_{2k} - \theta_{k,2k}} \\ D_{u,k} &= \frac{2(2 + \frac{(\sqrt{2}+1)\theta_{u,k}}{1-\delta_u})\theta_{u,2k}}{1 - \delta_{2k} - \theta_{k,2k}} \\ E_{u,k} &= \frac{\theta_{k,2k}}{1 - \delta_{2k} - \theta_{k,2k}}(2 + \frac{(\sqrt{2} + 1)\theta_{u,k}}{1 - \delta_u}) + 2(1 + \frac{\theta_{u,k}}{1 - \delta_u}) \\ e_0(T, \Delta) &= 2 \frac{\|x_{(T \cup \Delta)^c}\|_1}{\sqrt{|k|}} \end{aligned}$$

Proof: Proof is given in Appendix .

2.2 Regularized Modified-BPDN

In this section we introduce the regularized Modified CSDN and derive the bound for it's reconstruction error. like the previous section, we consider the case where there is one set of measurements y and a signal vector x .

$$y := Ax + w, \quad \text{where } \|w\| \leq \epsilon$$

Let N denote the support of x , i.e $N := \{i : x_i > 0\}$. Assume that we know partial part of support denoted by T . We define $\Delta := N \setminus T$. Also we assume partial knowledge of the signal estimate on T , and denote it by $\hat{\mu}_T$.

Regularized modified-BPDN solves the following problem.

$$\min_b \frac{1}{2} \|y - Ab\|_2^2 + \frac{1}{2} \lambda \|b_T - \hat{\mu}_T\|_2^2 + \gamma \|b_{T^c}\|_1 \quad (2.3)$$

In the following definition, we first define some variables that will be used repeatedly through the thesis.

Definition 3. *Let*

$$Q_T(S) = (A'_{T \cup S} A_{T \cup S} + \lambda \begin{pmatrix} I_T & 0_{T,S} \\ 0_{S,T} & 0_{S,S} \end{pmatrix}) \quad (2.4)$$

$$c_T(S) = Q_T(S)^{-1} (A'_{T \cup S} y + \lambda \hat{\mu}_{T \cup S}) \quad (2.5)$$

$$ERC(T, S, \lambda) = (1 - \max_{i \notin T \cup S} \|(A'_S M A_S)^{-1} A'_S M A_i\|_1) \quad (2.6)$$

$$M_{T,\lambda} \triangleq I - A_T (A'_T A_T + \lambda I_{|T|})^{-1} A'_T \quad (2.7)$$

Notice that for simplification, through the rest of report we use $Q(S)$, $c(S)$, $ERC(T, S)$, M instead of $Q_T(S)$, $c_T(S)$, $ERC(T, S, \lambda)$, $M_{T,\lambda}$. In the following lemma we bring the conditions under-which the problem (2.3) has a unique solution and we calculate the l_2 distance of this solution from the true signal.

Theorem 2. *Let $u = |T|$ and $k = |\Delta|$. If A_Δ has full rank, and :*

1. $ERC(u, k, \lambda) > 0$

2. $\gamma \geq \frac{\|(y - A c(\Delta))\|_2}{ERC(u, k, \lambda)}$

Then

1. *The function in (2.3) has a unique minimizer \hat{x} .*

2. *The error can be bounded by the following formula*

$$\|x - \hat{x}\|_2 \leq \gamma \sqrt{k} \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| + \|\lambda Q(\Delta)^{-1} (x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta})\| + \|Q(\Delta)^{-1} A'_T y\|$$

Proof: The proof has given in Appendix.

In the next lemma we rewrite the condition 1 and 2 of lemma 2 in terms of RIP and ROP constants. We do the same thing for (2.8).

Lemma 1. *If A_Δ has full rank, and*

1. $\lambda \geq \max(0, g_1(u, k), g_2(u, k))$
2. $\min(1 - \delta_k - \theta_{u,k}, 1 - \delta_k - \sqrt{k}\theta_{k,1}) \geq 0$
3. $\gamma \geq H_2(u, k, \lambda, \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\|)$

then

1. The function in (2.3) has a unique minimizer \hat{x}
2. $\|x - \hat{x}\|_2 \leq \gamma f_1 + f_2 \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + f_3 \|w\|$

where

$$\begin{aligned}
 g_1(u, k) &= \delta_u - 1 + \frac{\theta_{u,k}^2}{1 - \delta_k - \theta_{u,k}} \\
 g_2(u, k) &= -1 + \delta_u + \frac{\theta_{u,k}^2 + \sqrt{k}\theta_{u,k}\theta_{u,1}}{1 - \delta_k - \sqrt{k}\theta_{k,1}} \\
 H_2(u, k) &= \frac{\lambda \frac{\theta_{(u+k,1)}}{\min(1-\delta_u+\lambda, 1-\delta_k)-\theta_{u,k}} \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + (1 + \frac{\theta_{u+k,1}(\sqrt{1+\delta_{u+k}})}{\min(1-\delta_u+\lambda, 1-\delta_k)-\theta_{u,k}})\epsilon}{1 - \sqrt{k} \frac{\theta_{k,1} + \frac{\theta_{u,k}\theta_{u,1}}{1-\delta_u+\lambda}}{1 - \delta_k - \frac{\theta_{u,k}^2}{(1-\delta_u+\lambda)}}}
 \end{aligned}$$

and $f_i \equiv f_i(u, k, \lambda) \quad i = 1, 2, 3$

$$f_1(u, k, \lambda) = \sqrt{k} \sqrt{\frac{\theta_{u,k}^2}{(1 - \delta_u + \lambda)^2} + 1} \cdot \frac{1}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}} \quad (2.9)$$

$$f_2(u, k, \lambda) = \frac{\lambda}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \quad (2.10)$$

$$f_3(u, k, \lambda) = \frac{\sqrt{1 + \delta_{u+k}}}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \quad (2.11)$$

CHAPTER 3. Signal Models for Sparse and Compressible Sequential Signals

In this chapter, we bring two signal models that we will use for analysing our sequential algorithm in next chapter. Signal model I and signal model II have been defined in section 3.1 and 3.2. At section 3.3, we compare the advantage and artifacts of the both signal models.

3.1 Signal Model I

The proposed algorithm *does not* assume any signal model. But to prove its stability, we need certain assumptions on the signal changes over time. We use the Signal Model introduced in [26] as our signal sequence over time.

Assume the following

1. (addition) At each $t > 0$, S_a new coefficients get added to the support at magnitude r . Denote this set by A_t .
2. (increase) At each $t > 0$, the magnitude of S_a coefficients which had magnitude $(j - 1)r$ at $t - 1$ increases to jr . This occurs for all $2 \leq j \leq d$. Thus the maximum magnitude reached by any coefficient is $M := dr$.
3. (decrease) At each $t > 0$, the magnitude of S_a coefficients which had magnitude $(j + 1)r$ at $t - 1$ decreases to jr . This occurs for all $1 \leq j \leq (d - 1)$.
4. (removal) At each $t > 0$, S_a coefficients which had magnitude r at $t - 1$ get removed from the support (magnitude becomes zero). Denote this set by R_t .
5. (initial time) At $t = 0$, the support size is S_0 and it contains $2S_a$ elements each with magnitude $r, 2r, \dots, (d - 1)r$, and $(S_0 - (2d - 2)S_a)$ elements with magnitude M .

Notice that, in the above model, the size and composition of the support at any t is the same as that at $t = 0$. Also, at each t , there are S_a new additions and S_a removals. The new coefficient magnitudes increase gradually at rate r and do not increase beyond a maximum value $M := dr$. Similarly for decrease. The support size is always S_0 and the signal power is $(S_0 - (2d - 2)S_a)M^2 + 2S_a \sum_{j=1}^{d-1} j^2 r^2$. To understand the implications of the assumptions in Signal Model , we define the following sets.

Definition 4. *Let*

1. $D_t(j) := \{i : |x_{t,i}| = jr, |x_{t-1,i}| = (j+1)r\}$ *denote the set of elements that decrease from $(j+1)r$ to jr at time, t ,*
2. $I_t(j) := \{i : |x_{t,i}| = jr, |x_{t-1,i}| = (j-1)r\}$ *denote the set of elements that increase from $(j-1)r$ to jr at time, t ,*
3. $S_t(j) := \{i : 0 < |x_{t,i}| < jr\}$ *denote the set of small but nonzero elements, with smallness threshold jr .*
4. *Clearly,*
 - (a) *the newly added set, $A_t := I_t(1)$, and the newly removed set, $R_t := D_t(0)$.*
 - (b) *$|I_t(j)| = S_a$, $|D_t(j)| = S_a$ and $|S_t(j)| = 2(j-1)S_a$ for all j .*

Consider a $1 < j \leq d$. From the signal model, it is clear that at any time, t , S_a elements enter the small elements' set, $S_t(j)$, from the bottom (set A_t) and S_a enter from the top (set $D_t(j-1)$). Similarly S_a elements leave $S_t(j)$ from the bottom (set R_t) and S_a from the top (set $I_t(j)$). Thus,

$$S_t(j) = S_{t-1}(j) \cup (A_t \cup D_t(j-1)) \setminus (R_t \cup I_t(j)) \quad (3.1)$$

Since the sets $A_t, R_t, D_t(j-1), I_t(j)$ are mutually disjoint, and since $R_t \subseteq S_{t-1}(j)$ and $I_t(j) \subseteq S_{t-1}(j)$, thus,

$$S_{t-1}(j) \cup A_t \setminus R_t = S_t(j) \cup I_t(j) \setminus D_t(j-1) \quad (3.2)$$

3.2 Signal Model II

Inspired by [28], we propose a signal model in this part which basically relies on this fact that in some application, e.g. MRI imaging, the signal vector entries will have limited variations. We define k disjoint sets, $\Lambda_1 \dots \Lambda_k \subset [1..n]$ such that $\Lambda_1 \cup \Lambda_2 \cup \dots \cup \Lambda_k = [1..n]$. Let's define $t_0 \dots t_k \in \mathbb{R}$ such that $t_0 = 0 < t_1 < \dots < t_k$ and $c \in \mathbb{R}$. We establish $\mu \in \mathbb{R}^n$ as following

$$\mu_i \sim P_1 \times U\left(\frac{t_{j-1}}{1-c}, \frac{t_j}{1+c}\right) \quad i \in \Lambda_j$$

Where $0 \leq j \leq k$ and P_1 and U denote $1 - 2 \times \text{Bernoulli}(\frac{1}{2})$ and Uniform distribution.

At time $t=0$, we set $x_0 = \mu$. We build the signal, $x \in \mathbb{R}^n$, as follow

$$(x_i)_t = (x_i)_{t-1} + P_2 \times S_1 \times S_2 \times \frac{c}{s} \mu_i + P_3 \times S_3 \frac{c}{s} \mu_i$$

where $S_1 = \text{sgn}(\mu_i) \text{sgn}((x_i)_{t-1} - \mu_i)$, $S_2 = \text{sgn}(1 - \text{sgn}(|\frac{c}{s} \mu_i| - |(x_i)_{t-1} - \mu_i|))$, $S_3 = \text{sgn}(|\frac{c}{s} \mu_i| - |(x_i)_{t-1} - \mu_i|)$ and P_2, P_3 are random variables that return an integer from the sets $\{0, 1\}$ and $\{-1, 0, 1\}$ with the same probability.

Notice that S_2 and S_3 are two functions which return 0 and 1 if $(x_i)_{t-1} = \mu_i \pm \frac{c}{s} \mu_i$, respectively, otherwise they return 1 and 0.

3.2.1 Discussion of Signal Model

In this part, we focus on MRI Images and see how well the signal model can represent the MRI image signals. As an example, we take a sequence of MRI images of the cardiac system over 20 consecutive times. In Part (a) of Figure 1, the sorted form of the MRI coefficients in wavelet domain has been plotted.

If we look at the sorted values of coefficients we can see that these coefficients can be split in five categories based on their values that are independent of the time.

- A low fraction of large coefficients which constitute more than 0.90 of signal energy. In this case these are less than 0.001 of the whole number of coefficients.
- A fraction of middle range values which constitute more than 0.094 of signal energy. In this case there are less than 0.015 of the coefficients.

- A fraction of coefficients that have lower values and constitute more than .048 of signal energy. In this case these are less than 0.22 of the whole number of coefficients.
- A fraction of coefficients that have lower values and constitute about .001 of signal energy. In this case these are less than 0.01 of the whole number of coefficients.
- A large fraction of coefficients that have values near zero and constitute less than 0.001 of signal energy. In this case these are more than 0.77 of the whole number of coefficients.

If we track the changes of coefficients over time, we observe that the deviations are bounded by time independent values. This deviation is dependent on the value of the coefficient, e.g. larger coefficients have larger variations and vice versa. In the proposed signal model, we have considered this characterization and have bounded the variation of each coefficient. We use parameter c to do that. Therefore, for $t \neq t'$, it gives $|(x_i)_t - (x_i)_{t'}| \leq 2c|\mu_i|$.

Based on these observations we choose $k = 5$ in the signal model and set $|\Lambda_1| = 751$, $|\Lambda_2| = 40$, $|\Lambda_3| = 210$, $|\Lambda_4| = 15$, $|\Lambda_5| = 8$.

We set $t_1 = 8$, $t_2 = 22$, $t_3 = 100$, $t_4 = 300$ and $t_5 = 1400$ which define the range where the coefficients can change in each set.

In part (b) of Figure 1, we have plotted the sorted values of the sample signal which has been generated by the proposed signal model. Notice that by these values the real signal and proposed signal model energy are about 5.5×10^6 . Moreover, the generated signal follow close energy pattern of the MRI signals which was mentioned earlier. This is important in compressive sensing since the support size can be assumed as the lowest number of coefficients which accumulate a certain amount of energy.

Another fact which is important in the modeling is the variation of signal values over time. The parameter s control this variation in a way that for $i \in [1..n]$, we have $|(x_i)_t - (x_i)_{t-1}| = c/s|\mu_i|$. By setting $s = 20$, we would have about 0.03 variation in signal energy over consecutive times which is the same as the samples MRI sequence. Figure 1 shows the sorted coefficients of the generated signal. As it can be seen, the proposed signal model gives a closed model of MRI signal.

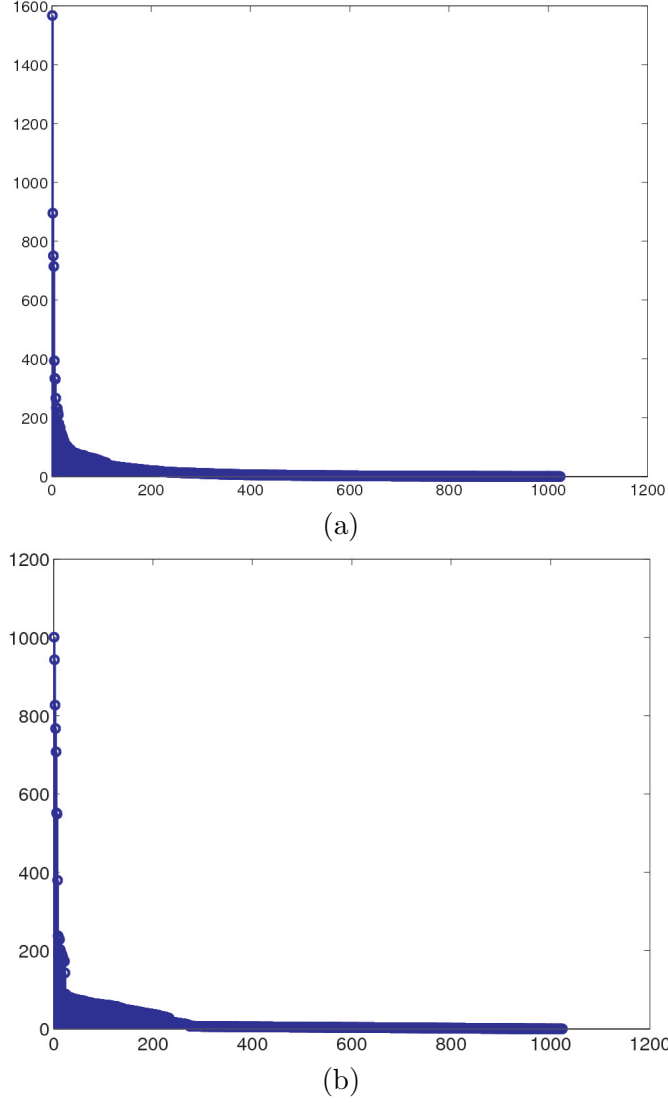


Figure 3.1 (a) Sorted MRI signal coefficient values in Wavelet domain (b) Sorted proposed signal model coefficient values in Wavelet domain

3.3 Comparison of signal model I and II

In sequential signal modeling, we are interested in a signal model which satisfy three characteristics

- It has the capability such that the support set changes dynamically over time.
- At each time instant, it generates a signal which its entries distribution follow the same distribution as the real signals.

- It contains the mechanism that the signal entries evolve over time.

The signal model I has been designed in a way that it gives the first and third conditions. There we saw that the signal entries are just limited to some certain steps that the distance between the consecutive levels, define the signal changes over time. So in this case, the evolution of signal over time is dependent on the signal energy at each time instant and it is not something that happens in applications.

The signal model II is successful with the last two items but the signal support does not change properly dynamically over time. This is actually true in MRI reconstruction but may not hold in some applications like video surveillance.

More precisely, if we want to have a signal which have the same energy over time, it should lie on the l_2 ball. Each signal vector is associated with one point on this ball and for the case where we are limited to certain distributions, these points will be on different parts of balls and have different distances. In other word, the evolution of the signal over time is not independent of the distribution of the signal at each time instant. So it is hard that a signal model satisfy all three items and we usually have a trade-off between those.

CHAPTER 4. Algorithm for Sequential Reg-mod-CSN and Stability Results

In this chapter, we bring algorithm which utilize reg-mod-csn and reg-mod-bpdn over time. Furthermore, we will analyze the performance of the two algorithms over time and obtain the conditions under which the reconstruction error is bounded by a time invariant value.

4.1 Algorithm for Sequential Regularized Modified CSN over time

Regularized Modified CSN was introduced in the previous section as the solution to the problem of (2.1). In other word, Regularized Modified CSN is the solution to the problem of sparse reconstruction (2.2) with partial knowledge of the support and signal value on the known support. For recursively reconstruction a time sequence of sparse signals, we use the support estimate from the previous time, \tilde{T}_{t-1} as the set T and use the signal estimate from the previous time on this support, $(\hat{x}_{t-1})_T$ as the μ_T . At the initial time, $t = 0$, we let T be the empty set, i.e we do simple CS. Therefore at $t = 0$ we need more measurements, $m_0 > m$. Denote the $m_0 \times n$ measurement matrix used at $t = 0$ by A_0 .

We summarize the Regularized modified CSN algorithm in Algorithm 1. Here α denote the support estimation threshold. Consider that in step 3 of algorithm we update our support estimation as \tilde{T}_t at time t .

4.2 Stability Results with Signal Model I

In this part we are finding the conditions under which the error bound for proposed algorithm remains bounded. For this purpose, we should develop the conditions for a certain set of large coefficients to definitely get detected and the elements of Δ_e to definitely get deleted.

Algorithm 1 Regularized Modified CSN over time

For $t \geq 0$, do

1. *Simple CS.* If $t = 0$, set $T_0 = \emptyset$ and compute \hat{x}_0 as the solution of

$$\min \|\beta\|_1 \quad \text{s.t.} \quad \|y_0 - A_0\beta\|_2 \leq \epsilon \quad (4.1)$$

2. *Regularized Modified CSN.* If $t > 0$, set $T_t = \tilde{T}_{t-1}$ and compute \hat{x}_t as the solution of

$$\min \|\beta_{T_t^c}\|_1 \quad \text{s.t.} \quad \|y - A\beta\|_2 \leq \epsilon \quad \text{and} \quad \|\beta_{T_t} - \mu_{T_t}\|_2 \leq \gamma \quad (4.2)$$

3. *Estimate the Support.* Compute \tilde{T}_t as

$$\tilde{T}_t = \{i \in [1, m] : |(\hat{x}_t)_i| > \alpha\} \quad (4.3)$$

4. Set $\mu = \hat{x}_t$. Output \hat{x}_t . Feedback μ and \tilde{T}_t .
-

In the following lemma we bring some simple facts that we use through the proof of Theorem 4.

Proposition 1. *In the third step of Algorithm 1 we have the following facts*

1. *An $i \in N_t$ will definitely get detected if $|x_i| > \alpha + \|x_t - \hat{x}_t\|$. This follows since $\|x_t - \hat{x}_t\| \geq \|x_t - \hat{x}_t\|_\infty \geq |x_t - \hat{x}_t|_i$*
2. *Similarly, all $i \in \tilde{\Delta}_{e,t}$ (the zero elements of \tilde{T}_t) will definitely not get detected if $\alpha \geq \|x_t - \hat{x}_t\|$. This is true since if $(x_t)_i = 0$ and $(\hat{x}_t)_i$ get detected as nonzero value $(\hat{x}_t)_i$, then $\alpha \leq \|(\hat{x}_t)_i - (x_t)_i\| \leq \|x_t - \hat{x}_t\|$ which is a contradiction with the assumption $\alpha \geq \|x_t - \hat{x}_t\|$.*

Proposition 2. *Under proposed Signal Model we have*

$$\|(x_t)_{T_t} - (\hat{x}_{t-1})_{T_t}\|_2 \leq \|x_{t-1} - \hat{x}_{t-1}\|_2 + \sqrt{2dS_a}r$$

Proof: Proof is straightforward from the fact that by Signal Model we have $\|x_t - x_{t-1}\|_2 \leq \sqrt{2dS_a}r$.

In the following theorem we bring the conditions that makes the error bounded by a time independent value.

Theorem 3 (Stability of Regularized Modified CSN over time). *Assume the Signal Model given above. For a d_0 such that $1 \leq d_0 \leq d$, set $S_1 = (2d_0 - 2)S_a$. If the following conditions hold*

$$1. \min(1 - \delta_{S_0}, 1 - \delta_{2S_1} - \theta_{S_1, 2S_1} - 2(2 + \frac{(\sqrt{2}+1)\theta_{S_0, S_1}}{1-\delta_{S_0}})\theta_{S_0, 2S_1}) > 0$$

$$2. \gamma = \frac{C_{S_0, S_1}\epsilon + \sqrt{2dS_a}r}{1 - D_{S_0, S_1}}$$

$$3. \alpha = C_{S_0, S_1}\epsilon + D_{S_0, S_1}\gamma$$

4. r satisfy

$$r \geq \frac{2C_{S_0, S_1}\epsilon}{d_0(1 - D_{S_0, S_1}) - 2D_{S_0, S_1}\sqrt{2dS_a}}$$

(it ensures that $d_0 r \geq 2 \times (C_{S_0, S_1}\epsilon + D_{S_0, S_1}\gamma)$)

5. n_0 is large enough so that

$$\|x_0 - \hat{x}_0\| \leq C_{S_0, S_1}\epsilon + D_{S_0, S_1}\gamma$$

Then we can conclude that

$$1. |T_t| \leq S_0, |\Delta_t| \leq S_1$$

$$2. \|x_t - \hat{x}_t\| \leq C_{S_0, S_1}\epsilon + D_{S_0, S_1}\gamma$$

Proof: Our approach for the proof is based on induction. Assume that the results hold at $t - 1$. Using condition 2 and Proposition 2, we can show that $\|x_t - \hat{x}_{t-1}\| \leq \gamma$. Condition 2 is meaningful when $D_{S_0, S_1} < 1$ which is equivalent to the second term of condition 1.

Next, we try to show that $|T_t| \leq S_0$ and $|\Delta_t| \leq S_1$. Finally, this, along with conditions 1 and 2 allows us to apply Theorem 1 to get the bound on $\|x_t - \hat{x}_{t-1}\|$. To show $|T_t| \leq S_0$ and $|\Delta_t| \leq S_1$, we first use the induction assumption, conditions 3 and 4 and Proposition 1 to bound $|\tilde{T}_{t-1}|$ and $|\tilde{\Delta}_{t-1}|$; and then use the signal model to bound $|T_t|$ and $|\Delta_t|$. The complete proof is given in the Appendix.

4.2.1 Discussion of Theorem

We can observe some results from Theorem 2. As we can see in the first condition of Theorem 2, reg-mod-CSN needs two requirements to hold, $\delta_{S_0} < 1$ and $1 - \delta_{2S_1} - \theta_{S_1, 2S_1} - \alpha > 0$

where $\alpha = 2(2 + \frac{(\sqrt{2}+1)\theta_{S_0,S_1}}{1-\delta_{S_0}})\theta_{S_0,2S_1}$. Consider the case where $\delta_{S_0} = \frac{3}{4}$ and $\theta_{S_0,2S_1} = \frac{1}{8}$ then it can be concluded that $\alpha = \frac{7}{8}$. So the second requirement of condition 1 is simplified to $\delta_{2S_1} + \theta_{S_1,2S_1} \leq \frac{1}{8}$. Since in practise S_1 is small in compare with S_0 , we can see that the condition $\delta_{2S_1} + \theta_{S_1,2S_1} \leq \frac{1}{8}$ will be satisfied easily.

We showed that if $\delta_{S_0} = \frac{3}{4}$ and $\theta_{S_0,2S_1} = \frac{1}{8}$ then Theorem 2. Comparing these with the results for modified-CS [26], $\delta_{S_0+S_1} \leq \frac{\sqrt{2}-1}{2}$, we observe that reg-mod-CSN remain stable under weaker conditions.

Also recall that CS results [5] needs $\delta_{2S_0} \leq \sqrt{2}-1$ that is an stronger condition in compare with $\delta_{S_0} = \frac{3}{4}$ which we obtained for reg-mod-CSN.

4.2.2 Simulation Result

We compared regularized modified CSN, modified BPDN [29], modified CS [30] and simple CS for different values of $\frac{S_0}{m}$. In Figure 1 we used Signal Model with $m = 100$, $n = 50$, $S_0 = 20, 30, 40$, $S_a = 1$, $r = \frac{1}{6}$ and $w_t \sim^{iid} \text{uniform}(-c, c)$ with $c = .05$. γ_1 is the value of γ in minimization problem (2.1) regularized modified CSN and γ_2 is the value of γ for problem (2) in [29] for modified BPDN respectively. The measurement matrix was random Gaussian. The simulation results have been obtained by averaging over 100 samples. We set the α to some value in the noise level($\alpha = .1$). By this value it gives a fairly accurate estimate of nonzero elements with a low number of falsely detections. In Figure 1 we showed a set of plots. Normalized MSE (NMSE), average number of extras(mean of the $|T_t \setminus N_t|$ over the 100 simulations) and average number of misses (mean of $|N_t \setminus T_t|$) are plotted in parts (b) and (c). Since in (b) and (c) the error was over 0.2 for CS, we just showed CS in (a). As it can be seen in (a) ($\frac{S_0}{m} = .2$) reg-mod-CSN ,mod-BPDN and mod-CS are stable and works almost the same (the errors are under 0.02) while the CS has a large error. In (b) as $\frac{S_0}{m}$ is increased ($\frac{S_0}{m} = .3$) modified BPDN and modified CS starts to become unstable(The NMSE is increased gradually over time) while reg mod CSN is still stable(The NMSE remains under 0.02 over time). In the case where $\frac{S_0}{m} = .4$ all three methods become unstable.

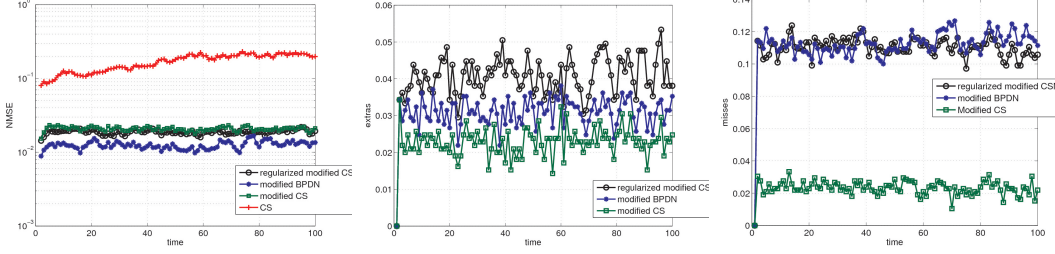
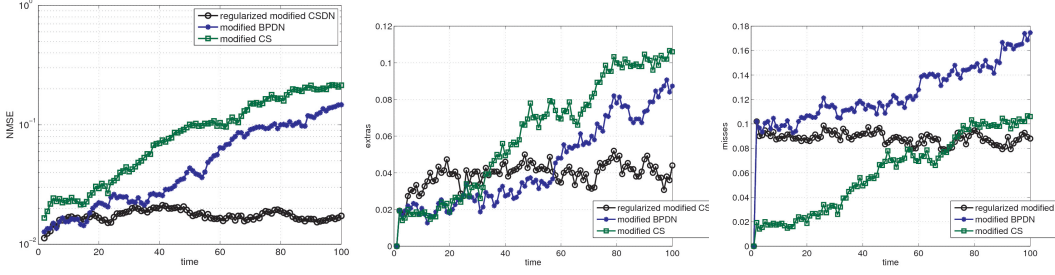
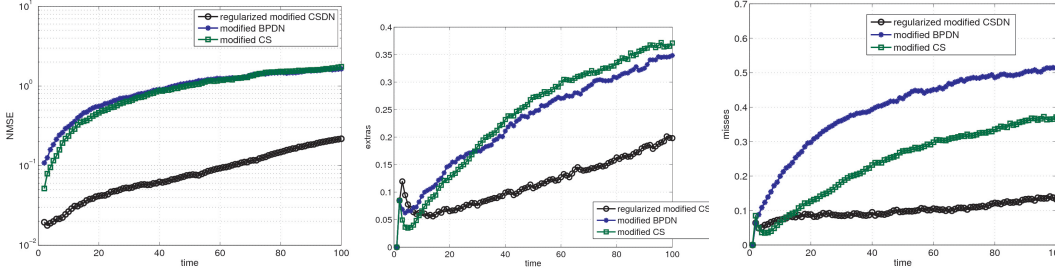
(a) $n = 50, m = 100, S_0 = 20, \gamma_1 = .4, \gamma_2 = .3$ (b) $n = 50, m = 100, S_0 = 30, \gamma_1 = .4, \gamma_2 = .3$ (c) $n = 50, m = 100, S_0 = 40, \gamma_1 = .4, \gamma_2 = .3$

Figure 4.1 Normalized MSE (NMSE), number of extras and number of misses over time for CS, modified CS, modified BPDN, and regularized modified CSN. In part (b) and (c), NMSE for CS was more than 20%.(plotted only in (a))

4.3 Stability Results with Signal Model II

Theorem 4 (Stability of Regularized Modified CSN over time). *Assume the Signal Model given above. IF there exists $1 \leq k_0 \leq k$ such that $S_0 = n - |\Lambda_1| - \dots - |\Lambda_{k_0-1}|$, $S_1 = |\Lambda_{k_0}|$ and the following conditions hold*

1. $\min(1 - \delta_{S_0}, 1 - \delta_{2S_1} - \theta_{S_1, 2S_1} - 2(2 + \frac{(\sqrt{2}+1)\theta_{S_0, S_1}}{1-\delta_{S_0}})\theta_{S_0, 2S_1}) > 0$
2. $\gamma = \frac{C_{S_0, S_1} \epsilon}{1 - D_{S_0, S_1}} + \frac{\sqrt{\frac{\epsilon}{s}} D_{S_0, S_1} \|\mu_{(\Lambda_1 \cup \dots \cup \Lambda_{k_0-1})^c}\|_2}{1 - D_{S_0, S_1}}$

- $$+ \frac{2E_{S_0, S_1}}{1-D_{S_0, S_1}} \frac{t_{k_0-1}|\Lambda_{k_0-1}|+\dots+t_1|\Lambda_1|}{|\Lambda_{k_0}|}$$
3. $\alpha = \frac{t_{k_0}+t_{k_0-1}}{2}$
 4. $\gamma \leq \frac{t_{k_0}-t_{k_0-1}}{2}$
 5. n_0 is large enough so that

$$\begin{aligned} \|x_0 - \hat{x}_0\| &\leq C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma \\ &+ 2E_{S_0, S_1} \frac{t_{k_0-1}|\Lambda_{k_0-1}|+\dots+t_1|\Lambda_1|}{|\Lambda_{k_0}|} \end{aligned}$$

Then we can conclude that

1. $|T_t| \leq S_0$, $|\Delta_t| \leq S_1$
2. $\|x_t - \hat{x}_t\| \leq C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma$
 $+ 2E_{S_0, S_1} \frac{t_{k_0-1}|\Lambda_{k_0-1}|+\dots+t_1|\Lambda_1|}{|\Lambda_{k_0}|}$

Proof: Our approach for the proof is based on induction. The condition 5 of the Theorem 2 gives the base case. For the induction step, assume that the results hold at $t-1$. By the third step of Algorithm 1, we know that

$$T_t = \tilde{T}_{t-1} = \{i \in [1, m] : |(\hat{x}_{t-1})_i| > \alpha\}$$

We define $\Delta_t = N_t \setminus T_t$ where

$$N_t = \{i \in [1, m] : |(x_{t-1})_i| > t_{k_0-1}\}$$

Note that based on signal model $N_t = N$ where $N = \Lambda_{k_0} \cup \dots \cup \Lambda_k$. By the choice of α in the Theorem 2 and the Proposition 1, we can conclude that

$$\Lambda_{k_0+1} \cup \dots \cup \Lambda_k \subset T_t \subset \Lambda_{k_0} \cup \dots \cup \Lambda_k$$

This suffice to conclude that $|T_t| \leq S_0$, $|\Delta_t| \leq S_1$. It remains to show the second conclusion of the Theorem 2. For that, we use inequality (2.2). By the already obtained bound for $|T_t|$ and $|\Delta_t|$, we set $u = S_0$ and $k = S_1$ and $\mu = \hat{x}_{t-1}$. The only thing that remains is to show that $\|(x_t)_{T_t} - (\hat{x}_t)_{T_{t-1}}\| \leq \gamma$. Notice that

$$\|(x_t)_{T_t} - (\hat{x}_t)_{T_{t-1}}\| \leq \|(x_t)_{T_t} - (x_{t-1})_{T_t}\| + \|x_{t-1} - \hat{x}_{t-1}\| \quad (4.4)$$

By using $\|(x_t)_{T_t} - (x_{t-1})_{T_t}\| \leq \sqrt{\frac{c}{s}} \mu_{(\Lambda_1 \cup \dots \cup \Lambda_{k_0-1})^c}$ and using the assumption of the induction for bounding $\|x_{t-1} - \hat{x}_{t-1}\|$ and employing the condition 2 of the Theorem 2, we can conclude that the right side of (4.4) is less than or equal to γ .

4.3.1 Discussion of Theorem

We can observe some results from Theorem 2. As we can see in the first condition of Theorem 2, reg-mod-CSN needs two requirements to hold, $\delta_{S_0} < 1$ and $1 - \delta_{2S_1} - \theta_{S_1, 2S_1} - \alpha > 0$ where $\alpha = 2(2 + \frac{(\sqrt{2}+1)\theta_{S_0, S_1}}{1-\delta_{S_0}})\theta_{S_0, 2S_1}$. Consider the case where $\delta_{S_0} = \frac{3}{4}$ and $\theta_{S_0, 2S_1} = \frac{1}{8}$ then it can be concluded that $\alpha = \frac{7}{8}$. So the second requirement of condition 1 is simplified to $\delta_{2S_1} + \theta_{S_1, 2S_1} \leq \frac{1}{8}$. Since in practice S_1 is small in compare with S_0 , we can see that the condition $\delta_{2S_1} + \theta_{S_1, 2S_1} \leq \frac{1}{8}$ will be satisfied easily.

We showed that if $\delta_{S_0} = \frac{3}{4}$ and $\theta_{S_0, 2S_1} = \frac{1}{8}$ then Theorem 2. Comparing these with the results for modified-CS [26], $\delta_{S_0+S_1} \leq \frac{\sqrt{2}-1}{2}$, we observe that reg-mod-CSN remain stable under weaker conditions.

Also recall that CS results [5] needs $\delta_{2S_0} \leq \sqrt{2} - 1$ that is an stronger condition in compare with $\delta_{S_0} = \frac{3}{4}$ which we obtained for reg-mod-CSN.

4.3.2 Simulation Result

In this section we compare the reconstruction error of regularized modified CSN, modified CS [16] and simple CS. The comparison has been made for different values of m .

MRI system generate the 2D Fourier of the object as measurements

$$Y = FOF' \quad (4.5)$$

where O represents the object, Y represents measurement matrix and F is the 2D Fourier matrix.

Moreover, object O shows high sparsity in wavelet domain so we pick the wavelet domain as our focusing domain

$$X = WOW' \quad (4.6)$$

where W denotes 2D wavelet matrix. Using this, we can rewrite (4.5) as

$$Y = FWXW'F' \quad (4.7)$$

By stacking Y and X into vectors y and x and setting $A = (F \otimes F')(W \otimes W')$ we would have

$$y = Ax$$

which is the 1D version of the equation (4.7).

We used the proposed signal model with the same values of parameters as section 3.2.1 and different values of m . we set $k_0 = 2$. We chose the support as $N_t = \{i \in [1, m] : |(x_{t-1})_i| > 16\}$ which on average contains more than 99% of energy.

We run the algorithm for three different values of $m=400,500,600$. The parameter α is chosen based on the number of measurements. Notice that based on Theorem , the error bound depends on ϵ , γ , $|T|$ and the signal residual $x_{(T \cup \Delta)^c}$. This gives a key for choosing α . Notice that for low values of α , we may count a lot of entries that leads to a large value of $|T|$ and lower value for $x_{(T \cup \Delta)^c}$. Notice that large $|T|$ makes the RIP constants get larger values and consequently larger errors. So basically the choice of α is a tradeoff between $|T|$ and $x_{(T \cup \Delta)^c}$. We are not interested to count the small value entries as the support, $-T-$, since $x_{(T \cup \Delta)^c}$ gets low values for these entries. As we increase the number of measurements, we would have larger values for RIP constants that increase the error bound. By the results of the experiments, it is suggested that ones pick a larger α . Notice that $|T|$ is updated based on the reconstruction error from the previous time. Lower measurements means higher error, in other words, the previous time reconstruction are less reliable and by choosing low value for α , we may detects so many extra coefficients that actually does not contribute much to the signal energy. This can be amplified over time and result in an unstable system (large error). Based on the experiments, it is suggested that we chose larger values for α as m decrease.

For regularized modified CSN, α is set to 35, 35, 30 for $m = 400, 500, 600$ and for modified CS, α is set to 50, 50, 40 accordingly.

Noise is generated as $w_t \sim^{iid} \text{uniform}(-5, 5)$. Condition 2 of the Theorem 2 gives us an approximation of how to choose γ . Notice that large γ may gives and worthless constraint

that turn the problem to the modified CS. From the other side, low values of γ may give an constraint that does not hold in reality and doesn't let the algorithm to decide based on the measurements. Through our experiments we choose $\gamma = 20$.

In Figure 2, Normalized MSE (NMSE), average number of extras(mean of the $|T_t \setminus N_t|$ over the 5 simulations) and average number of misses (mean of $|N_t \setminus T_t|$) are plotted in parts (a), (b) and (c).

The simulation results have been obtained by averaging over 50 samples. As it can be seen in (c) ($m = 600$) reg-mod-CSN and mod-CS are stable and the errors are less than 0.01 while the CS has a large error. In (b) as m is decreased ($m = 400$) all three methods get larger error while reg-mod-CSN works the best followed by mod-CS and CS. In the case where $m = 300$ all three methods all three methods get large errors and have close performance.

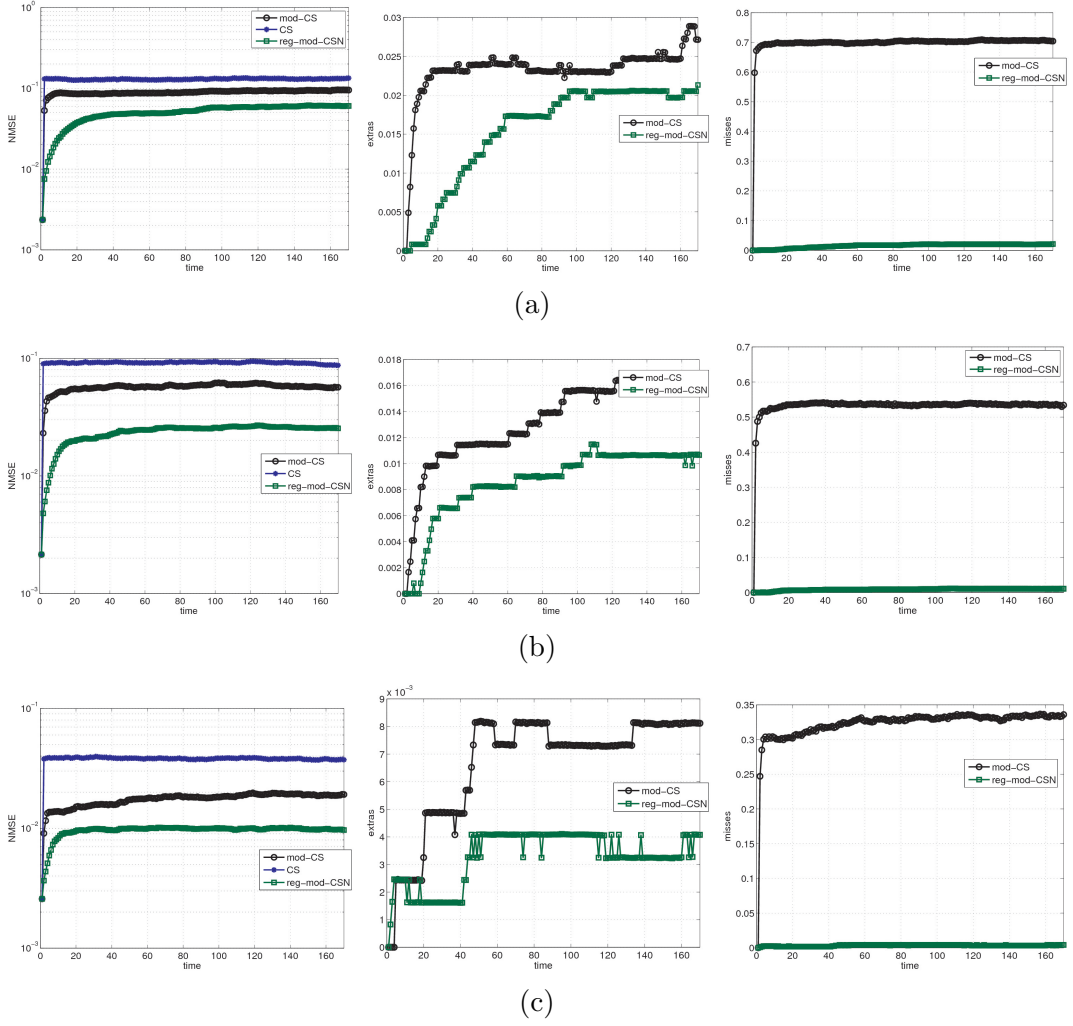


Figure 4.2 Normalized MSE (NMSE), number of extras and number of misses over time for CS, modified CS, modified BPDN, and regularized modified CSN generated signal by signal model. In part (b) and (c), NMSE for CS was more than 20%.(plotted only in (a))

CHAPTER 5. Summary and Future Work

In this work, we focused on the problem of sparse reconstruction via partial knowledge of support and an erroneous signal value estimate on this support. We proposed two optimization problem (reg-mod-CSN and reg-mod-BPDN) which is based on the l_1 minimization proposed in Compressive Sensing. We found the conditions which these two problems give an unique solution and we found the error bound for each. We particularly proposed an algorithm using reg-mod-CSN in sequential reconstruction in a way that the initial support and signal value estimate is provided by the previous time estimate. Considering the stability of this algorithm, we brought two signal models which represent the sequential signals over time. We developed conditions which the system remain stable with these two signal models. we also discussed the weak and strong point of the signal model and its impact on stability results. We also run the algorithm with some random sample to test the efficiency of our method In the simulation part, it was shown that using this algorithm leads to lower error bound with lower number of measurements in compare with the former methods, e.g. mod-CS and CS.

APPENDIX A. APPENDIX

A.1 Proof of Theorem 1

Here we consider the general case where the signal is not sparse. Let $V = [1...m]$. Assume that we know partial part of support denoted by T . We redefine set N such that $T \subset N \subset V$. To prove the Theorem, first let us get the following relation by using the fact that both x and \hat{x} are feasible

$$\|A(\hat{x} - x)\|_2 \leq \|A\hat{x} - y\|_2 + \|y - Ax\|_2 \leq 2\epsilon \quad (\text{A.1})$$

Basically our approach is a modification of the proof [5]. Let us write $\hat{x} = x + h$. Our aim in the rest of the proof is to make an upper bound for $\|h\|_2$. We decompose the vector h into a sum of vectors. We define $\Delta_0 = N \setminus T$ and Δ_j for $j \geq 1$ as the support of k largest coefficient of $h_{S_j^c}$ with $S_j = T \cup \bigcup_{l=0}^{j-1} \Delta_l$. The plan of the proof is to bound $\|h_T\|_2$, $\|h_{\Delta_0 \cup \Delta_1}\|_2$ and $\|h_{(T \cup \Delta_0 \cup \Delta_1)^c}\|_2$.

Using the triangular inequality, we have $\|h_{(T \cup \Delta_0 \cup \Delta_1)^c}\|_2 \leq \sum_{j=2} \|h_{\Delta_j}\|_2$. For $j \geq 1$, $\|h_{\Delta_j}\|_2 \leq k^{\frac{1}{2}} \|h_{\Delta_j}\|_\infty \leq k^{-\frac{1}{2}} \|h_{\Delta_{j-1}}\|_1$ this leads to

$$\|h_{(T \cup \Delta_0 \cup \Delta_1)^c}\|_2 \leq \sum_{j=2} \|h_{\Delta_j}\|_2 \leq \frac{1}{\sqrt{k}} \|h_{(T \cup \Delta_0)^c}\|_1. \quad (\text{A.2})$$

Since $\hat{x} = x + h$ is the solution to (2.1) and both \hat{x} and x are feasible, we have

$$\begin{aligned} \|x_{T^c}\|_1 &\geq \|(x + h)_{T^c}\|_1 \\ &= \|x_{\Delta_0} + h_{\Delta_0}\|_1 + \|x_{(T \cup \Delta_0)^c} + h_{(T \cup \Delta_0)^c}\|_1 \\ &\geq \|x_{\Delta_0}\|_1 - \|h_{\Delta_0}\|_1 + \|h_{(T \cup \Delta_0)^c}\|_1 - \|x_{(T \cup \Delta_0)^c}\|_1 \end{aligned}$$

So then we have

$$\|h_{(T \cup \Delta_0)^c}\|_1 \leq \|h_{\Delta_0}\|_1 + 2\|x_{(T \cup \Delta_0)^c}\|_1 \quad (\text{A.3})$$

First we bound $\|h_T\|_2$. To do that, observe that $Ah_T = Ah - \sum_{j=0} Ah_{\Delta_j}$ and, therefore,

$$\|Ah_T\|^2 = \langle Ah_T, Ah \rangle - \langle Ah_T, \sum_{j=0} Ah_{\Delta_j} \rangle$$

Applying Cauchy-Schwartz, it follows from (A.1) and the restricted isometry and orthogonality property that

$$\begin{aligned} (1 - \delta_u) \|h_T\|_2^2 &\leq \sqrt{1 + \delta_u} \|h_T\|_2 (2\epsilon) + \theta_{u,k} \|h_T\|_2 (\sum_{j=0} \|h_{\Delta_j}\|_2) \\ \|h_T\|_2 &\leq \frac{\sqrt{1 + \delta_u}}{1 - \delta_u} (2\epsilon) + \frac{\theta_{u,k}}{1 - \delta_u} (\sum_{j=0} \|h_{\Delta_j}\|_2) \end{aligned} \quad (\text{A.4})$$

We can break the term $\sum_{j=0} \|h_{\Delta_j}\|_2 = \|h_{\Delta_0}\|_2 + \|h_{\Delta_1}\|_2 + \sum_{j=2} \|h_{\Delta_j}\|_2$. Since $\|h_{\Delta_0}\|_2 + \|h_{\Delta_1}\|_2 \leq \sqrt{2} \|h_{\Delta_0 \cup \Delta_1}\|_2$, Using (A.2) we can conclude

$$\sum_{j=0} \|h_{\Delta_j}\|_2 \leq \sqrt{2} \|h_{\Delta_0 \cup \Delta_1}\|_2 + \frac{1}{\sqrt{k}} \|h_{(T \cup \Delta_0)^c}\|_1 \quad (\text{A.5})$$

Using (A.5) we can rewrite inequality (A.4) as,

$$\|h_T\|_2 \leq \frac{\sqrt{1 + \delta_u}}{1 - \delta_u} (2\epsilon) + \frac{\theta_{u,k}}{1 - \delta_u} (\sqrt{2} \|h_{\Delta_0 \cup \Delta_1}\|_2 + \frac{1}{\sqrt{k}} \|h_{(T \cup \Delta_0)^c}\|_1) \quad (\text{A.6})$$

In the next step we bound $\|h_{\Delta_0 \cup \Delta_1}\|_2$. To do that we first make a bound for $\|h_T\|_2$. Since both x and \hat{x} are feasible and by using the second constraint of problem (2.1) we have

$$\|h_T\|_2 = \|x_T - \hat{x}_T\|_2 \leq \|x_T - \mu_T\|_2 + \|\hat{x}_T - \mu_T\|_2 \leq 2\gamma \quad (\text{A.7})$$

Same as previous step we can write $\|Ah_{\Delta_0 \cup \Delta_1}\|_2^2$ as

$$\begin{aligned} \|Ah_{\Delta_0 \cup \Delta_1}\|_2^2 &= \langle Ah_{\Delta_0 \cup \Delta_1}, Ah \rangle - \langle Ah_{\Delta_0 \cup \Delta_1}, Ah_T \rangle - \\ &\quad \langle Ah_{\Delta_0 \cup \Delta_1}, A(\sum_{j=2} h_{\Delta_j}) \rangle \end{aligned}$$

Using Cauchy-Schwartz, (A.1) and the restricted isometry property we have

$$\langle Ah_{\Delta_0 \cup \Delta_1}, Ah \rangle \leq 2\sqrt{1 + \delta_{2k}\epsilon} \|h_{\Delta_0 \cup \Delta_1}\|_2 \quad (\text{A.8})$$

Employing the restricted orthogonally property and (A.7) we get

$$\langle Ah_{\Delta_0 \cup \Delta_1}, Ah_T \rangle \leq 2\theta_{u,2k}\gamma \|h_{\Delta_0 \cup \Delta_1}\|_2 \quad (\text{A.9})$$

Using the restricted orthogonally property we get

$$\langle Ah_{\Delta_0 \cup \Delta_1}, A\Sigma_{j=2}h_{\Delta_j} \rangle \leq \theta_{k,2k} \|h_{\Delta_0 \cup \Delta_1}\|_2 (\Sigma_{j=2} \|h_{\Delta_j}\|_2) \quad (\text{A.10})$$

Using (A.2) we can rewrite the above inequality as

$$\langle Ah_{\Delta_0 \cup \Delta_1}, A\Sigma_{j=2}h_{\Delta_j} \rangle \leq \frac{\theta_{k,2k}}{\sqrt{k}} \|h_{\Delta_0 \cup \Delta_1}\|_2 \|h_{(T \cup \Delta_0)^c}\|_1 \quad (\text{A.11})$$

Combining (A.8), (A.9) and (A.11), we get

$$\begin{aligned} (1 - \delta_{2k}) \|h_{\Delta_0 \cup \Delta_1}\|_2^2 &\leq \sqrt{1 + \delta_{2k}} (2\epsilon) \|h_{\Delta_0 \cup \Delta_1}\|_2 \\ &+ \theta_{u,2k} \|h_T\|_2 \|h_{\Delta_0 \cup \Delta_1}\|_2 + \frac{\theta_{k,2k}}{\sqrt{k}} \|h_{\Delta_0 \cup \Delta_1}\|_2 \|h_{(T \cup \Delta_0)^c}\|_1 \end{aligned} \quad (\text{A.12})$$

By simplifying the above inequality we have

$$\|h_{\Delta_0 \cup \Delta_1}\|_2 \leq 2 \frac{\sqrt{1 + \delta_{2k}}}{1 - \delta_{2k}} \epsilon + 2 \frac{\theta_{u,2k}}{1 - \delta_{2k}} \gamma + \frac{\theta_{k,2k}}{\sqrt{k}(1 - \delta_{2k})} \|h_{(T \cup \Delta_0)^c}\|_1 \quad (\text{A.13})$$

By inequality (A.3) we can conclude

$$\begin{aligned} \|h_{(T \cup \Delta_0)^c}\|_1 &\leq \sqrt{k} \|h_{\Delta_0}\|_2 + 2 \|x_{(T \cup \Delta_0)^c}\|_1 \\ &\leq \sqrt{k} \|h_{\Delta_0 \cup \Delta_1}\|_2 + 2 \|x_{(T \cup \Delta_0)^c}\|_1 \end{aligned} \quad (\text{A.14})$$

We use this inequality to replace it with $\|h_{(T \cup \Delta_0)^c}\|_1$ in inequality (A.13).

$$\begin{aligned} \|h_{\Delta_0 \cup \Delta_1}\|_2 &\leq 2 \frac{\sqrt{1 + \delta_{2k}}}{1 - \delta_{2k}} \epsilon + 2 \frac{\theta_{u,2k}}{1 - \delta_{2k}} \gamma \\ &+ \frac{\theta_{k,2k}}{1 - \delta_{2k}} \|h_{\Delta_0 \cup \Delta_1}\|_2 + 2 \frac{\theta_{k,2k}}{\sqrt{k}(1 - \delta_{2k})} \|x_{(T \cup \Delta_0)^c}\|_1 \end{aligned}$$

Simplifying the above inequality lead to

$$\|h_{\Delta_0 \cup \Delta_1}\|_2 \leq \tilde{F}_1 \epsilon + \tilde{F}_2 \gamma + \tilde{F}_3 e_0(T, \Delta_0) \quad (\text{A.15})$$

where

$$\begin{aligned} e_0(T, \Delta) &= 2 \frac{\|x_{(T \cup \Delta)^c}\|_1}{\sqrt{|\Delta|}} \\ \tilde{F}_1 &= \frac{2\sqrt{1 + \delta_{2k}}}{1 - \delta_{2k} - \theta_{k,2k}}, \tilde{F}_2 = \frac{2\theta_{u,2k}}{1 - \delta_{2k} - \theta_{k,2k}} \\ \tilde{F}_3 &= \frac{\theta_{k,2k}}{1 - \delta_{2k} - \theta_{k,2k}} \end{aligned}$$

Here we use the previous bounds on $\|h_T\|_2$, $\|h_{\Delta_0 \cup \Delta_1}\|_2$ and $\|h_{(T \cup \Delta_0 \cup \Delta_1)^c}\|_2$ to bound $\|h\|_2$. Using (A.6), (A.13) and (A.2) we have

$$\begin{aligned} \|h\| &\leq \|h_T\|_2 + \|h_{\Delta_0 \cup \Delta_1}\|_2 + \|h_{(T \cup \Delta_0 \cup \Delta_1)^c}\|_2 \leq \\ &\frac{\sqrt{1+\delta_u}}{1-\delta_u}(2\epsilon) + \frac{\theta_{u,k}}{1-\delta_u}(\sqrt{2}\|h_{\Delta_0 \cup \Delta_1}\|_2 \\ &+ \frac{1}{\sqrt{k}}\|h_{(T \cup \Delta_0)^c}\|_1) + \|h_{\Delta_0 \cup \Delta_1}\|_2 + \frac{1}{\sqrt{k}}\|h_{(T \cup \Delta_0)^c}\|_1 \end{aligned}$$

Using (A.14) and reordering the terms lead to

$$\begin{aligned} \|h\| &\leq 2\frac{\sqrt{1+\delta_u}}{1-\delta_u}\epsilon + (2 + \frac{(\sqrt{2}+1)\theta_{u,k}}{1-\delta_u})\|h_{\Delta_0 \cup \Delta_1}\|_2 \\ &+ 2(1 + \frac{\theta_{u,k}}{1-\delta_u})\frac{\|x_{(T \cup \Delta)^c}\|_1}{\sqrt{k}} \end{aligned}$$

By substitution of (A.15) in above inequality we get

$$\|h\| \leq C_{u,k}\epsilon + D_{u,k}\gamma + E_{u,k}e_0(T, \Delta_0)$$

where

$$E_{u,k} = \frac{\theta_{k,2k}}{1-\delta_{2k}-\theta_{k,2k}}(2 + \frac{(\sqrt{2}+1)\theta_{u,k}}{1-\delta_u}) + 2(1 + \frac{\theta_{u,k}}{1-\delta_u})$$

A.2 Proof of Lemma 2

In this section we study minimizing the function $L(b)$:

$$L(b) = \frac{1}{2} \|y - Ab\|^2 + \frac{1}{2}\lambda \|b_T - \hat{\mu}_T\|^2 + \gamma \|b_{T^c}\|_1 \quad (\text{A.16})$$

We are searching the sufficient conditions under which $L(b)$ has a unique minimizer in a way that the unique minimizer is supported on set $T \cup \Delta$. For achieving this goal, first we characterize $L(b)$ when it is restricted to coefficient vectors supported on $T \cup \Delta$. Then we find the new conditions that every perturbation away from the restricted minimizer increase the value of the objective function.

We characterize $L(b)$ over all coefficient vectors supported on $T \cup \Delta$ by the function $F(b)$:

$$F(b) = \frac{1}{2} \|y - A_{T \cup \Delta} b_{T \cup \Delta}\|^2 + \frac{1}{2}\lambda \|b_T - \hat{\mu}_T\|^2 + \gamma \|b_{T^c}\|_1$$

Since $F(b)$ is a proper convex function then b_* is a unique minimizer of $F(b)$ if and only if $0 \in \frac{\delta F(b)}{\delta b}$. Hence,

$$(A'_{T \cup \Delta} A_{T \cup \Delta})b_* - (A'_{T \cup \Delta} y) + \lambda \begin{bmatrix} I_{T,T} & 0_{T,\Delta} \\ 0_{\Delta,T} & 0_{\Delta,\Delta} \end{bmatrix} b_* - \lambda \begin{bmatrix} \hat{\mu}_T \\ 0_{\Delta,1} \end{bmatrix} + \gamma \begin{bmatrix} 0_{T,1} \\ g_\Delta \end{bmatrix} = 0$$

$$b_* = Q(\Delta)^{-1}(A'_{T \cup \Delta} y + \lambda \hat{\mu}_{T \cup \Delta} - \gamma g_{T \cup \Delta}) \quad (\text{A.17})$$

We should develop conditions which ensure that

$$L(b_* + h) - L(b_*) \geq 0 \quad (\text{A.18})$$

where h is a perturbation. Each perturbation admits a unique decomposition

$$h = u + v$$

We expand (A.18) to obtain

$$\begin{aligned} L(b_* + h) - L(b_*) = & \\ \frac{1}{2}(\|y - A(b_* + u) - Av\|^2 - \|y - Ab_*\|^2) + \frac{1}{2}\lambda(\|(b_* + u)_T - \hat{\mu}_T\|^2 - \|(b_*)_T - \hat{\mu}_T\|^2) \\ & + \gamma(\|(b_* + u)_{T^c} + v\|_1 - \|(b_*)_{T^c}\|_1) \end{aligned}$$

After some simplification, it gives

$$L(b_* + h) - L(b_*) = L(b_* + u) - L(b_*) + \frac{1}{2}\|Av\|_2^2 - \text{Re}\langle y - Ab_*, Av \rangle + \text{Re}\langle Au, Av \rangle + \gamma\|v\|_1$$

Since b_* is a unique minimizer over the set $T \cup \Delta$, therefore $L(b_* + u) - L(b_*) \geq 0$. This implies that for having $L(b_* + h) - L(b_*) \geq 0$ to be satisfied, it is sufficient to have

$$\gamma\|v\|_1 - |\langle y - Ab_*, Av \rangle| - |\langle Au, Av \rangle| \geq 0 \quad (\text{A.19})$$

Let's focus on the second term on the left side of inequality (A.19). We can write v as

$$v = [\Sigma_{\omega \notin T \cup \Delta} \theta_\omega e_\omega] \|v\|_1$$

Where $\|\theta\|_1 = 1$. By using the above equation we have

$$Av = [\Sigma_{\omega \notin T \cup \Delta} \theta_\omega A_\omega] \|v\|_1$$

Using the triangle inequality and then Jensen's inequality we obtain

$$|\langle y - Ab_*, Av \rangle| \leq [\Sigma_{\omega \notin T \cup \Delta} |\theta_\omega| |\langle y - Ab_*, A_\omega \rangle|] \|v\|_1 \leq \max_{\omega \notin T \cup \Delta} |\langle y - Ab_*, A_\omega \rangle| \cdot \|v\|_1$$

To control the third term of inequality (A.19), we use the standard operator norm

$$|\langle Au, Av \rangle| = |\langle A^* Au, v \rangle| \leq \|A^* Au\|_\infty \|v\|_1 \leq \delta \|A^* A\|_{\infty, \infty} \|v\|_1$$

Where we have used $\|u\|_\infty \leq \delta$. By applying these modification in second and third term of inequality (A.19), we can rewrite (A.19) as

$$[\gamma - \max_{\omega \notin T \cup \Delta} |\langle y - Ab_*, A_\omega \rangle| - \delta \|A^* A\|_{\infty, \infty}] \|v\|_1 \geq 0.$$

We can select δ as small as we want so the first two term of left-hand side is strictly positive for each small perturbation h , We can conclude that for having $L(b_* + h) - L(b_*) \geq 0$, it is sufficient to have

$$\gamma - |\langle y - Ab_*, A_i \rangle| > 0 \quad (\text{A.20})$$

Here we expand $|\langle y - Ab_*, A_i \rangle|$ by substituting b_* from (A.17) in above inequality

$$|\langle y - A_{T \cup \Delta} b_*, A_i \rangle| = |\langle A_{T \cup \Delta} Q(\Delta)^{-1} (\lambda \begin{pmatrix} x_T - \hat{\mu}_T \\ 0_\Delta \end{pmatrix} + \gamma \begin{pmatrix} 0_T \\ g_\Delta \end{pmatrix}) + (I - A_{T \cup \Delta} Q(\Delta)^{-1} A'_{T \cup \Delta}) w, A_i \rangle|$$

Using this we can rewrite (A.20) as

$$\gamma(1 - |\langle A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}, A_i \rangle|) \geq |\langle A_{T \cup \Delta} Q(\Delta)^{-1} \lambda(x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}) + (I - A_{T \cup \Delta} Q(\Delta)^{-1} A'_{T \cup \Delta}) w, A_i \rangle| \quad (\text{A.21})$$

Next, we focus on the left parantese in inequaity (A.21) to bring it in a simpler form.

First, we rewrite $A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}$ by substituing $A_{T \cup \Delta}$, $Q(\Delta)$ and $g_{T \cup \Delta}$ with their correspondent matrix

$$A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta} = (A_T A_\Delta) \begin{pmatrix} A'_T A_T + \lambda I_T & A'_T A_\Delta \\ A'_\Delta A_T & A'_\Delta A_\Delta \end{pmatrix}^{-1} \begin{pmatrix} 0_{T,1} \\ g_\Delta \end{pmatrix}$$

By using block matrix inversion and multiplying the second and third paranteses above we have

$$A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta} = (A_T A_\Delta) \begin{pmatrix} -(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta \\ (A'_\Delta M A_\Delta)^{-1} g_\Delta \end{pmatrix}$$

By multiplying the two remained parantheses we obtain

$$\begin{aligned} A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta} &= -A_T (A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta + A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta \\ &= M A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta \end{aligned} \quad (\text{A.22})$$

We use (A.22) to bound the term $|\langle A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}, A_i \rangle|$ in the left side of inequality (A.21)

$$|\langle A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}, A_i \rangle| = |\langle (M A_\Delta (A'_\Delta M A_\Delta)^{-1})' A_i, g_\Delta \rangle| = |\langle (A'_\Delta M A_\Delta)^{-1} A'_\Delta M' A_i, g_\Delta \rangle|$$

since $\|g_\Delta\|_\infty \leq 1$ we get

$$|\langle A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}, A_i \rangle| \leq \|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1$$

Therefore, we can conclude that the left paranthese in inequality (A.21) is less than $ERC(T, \Delta)$ where

$$ERC(T, \Delta) = (1 - \max_{i \notin T \cup \Delta} \|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1)$$

Notice the fact that the (A.21) become worthless when the left paranthese becomes less than zero. So having $ERC(T, \Delta) \geq 0$ ensures that $(1 - |\langle A_{T \cup \Delta} Q(\Delta)^{-1} g_{T \cup \Delta}, A_i \rangle|)$ is positive.

Now we consider the right side of (A.21). First, notice that

$$y - Ac(\Delta) = A_{T \cup \Delta} Q(\Delta)^{-1} \lambda (x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}) + (I - A_{T \cup \Delta} Q(\Delta)^{-1} A'_{T \cup \Delta}) w$$

By the assumption that $\|A_i\| \leq 1$ and using cauchy-shwartz we have $\langle y - Ac(\Delta), A_i \rangle \leq \|y - Ac(\Delta)\|$. Hence, we can conclude that if

$$\gamma \geq \frac{\|(y - Ac(\Delta))\|_2}{ERC(T, \Delta)} \quad (\text{A.23})$$

then the inequality (A.21) will be held. Thus the problem (2.3) would have a unique solution on the set $T \cup \Delta$ if

$$\gamma \geq \frac{\|(y - Ac(\Delta))\|_2}{ERC(T, \Delta)}$$

and

$$ERC(T, \Delta) > 0$$

We already showed the conditions which are necessary for having unique solution on the set $T \cup \Delta$. In the rest of this section we are going to find the error bound, $\|x - \hat{x}\|$, for the problem (2.3). From (A.17) we know the solution to the problem (2.3) is

$$b_* = Q(\Delta)^{-1}(A'_{T \cup \Delta} y + \lambda \hat{\mu}_{T \cup \Delta} - \gamma g_{T \cup \Delta})$$

Thus, we have

$$\begin{aligned} \|x - b_*\| &= \|Q(\Delta)^{-1}(\gamma g_{T \cup \Delta} + \lambda(x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}) - A'_{T \cup \Delta} w)\| \\ &\leq \|\gamma Q(\Delta)^{-1} g_{T \cup \Delta}\| + \|\lambda Q(\Delta)^{-1}(x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta})\| + \|Q(\Delta)^{-1} A'_{T \cup \Delta} w\| \quad (\text{A.24}) \\ &\leq \gamma \|Q(\Delta)^{-1} g_{T \cup \Delta}\| + \lambda \|Q(\Delta)^{-1}\| \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + \|Q(\Delta)^{-1}\| \|A'_{T \cup \Delta}\| \|w\| \end{aligned}$$

First we bound the term $\|Q(\Delta)^{-1} g_{T \cup \Delta}\|_2$. By substituting $Q(\Delta)$ and $g_{T \cup \Delta}$ with their corresponding matrix we obtain

$$Q(\Delta)^{-1} g_{T \cup \Delta} = \begin{pmatrix} A'_T A_T + \lambda I_T & A'_T A_\Delta \\ A'_\Delta A_T & A'_\Delta A_\Delta \end{pmatrix}^{-1} \begin{pmatrix} 0_{T,1} \\ g_\Delta \end{pmatrix}$$

Using block matrix inversion and multiplying the first and second parantheses above leads to

$$Q(\Delta)^{-1} g_{T \cup \Delta} = \begin{pmatrix} -(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta \\ (A'_\Delta M A_\Delta)^{-1} g_\Delta \end{pmatrix}$$

Hence, we have

$$\begin{aligned} \|Q(\Delta)^{-1} g_{T \cup \Delta}\| &= \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta (A'_\Delta M A_\Delta)^{-1} g_\Delta\|_2^2 + \|(A'_\Delta M A_\Delta)^{-1} g_\Delta\|^2} \\ \|Q(\Delta)^{-1} g_{T \cup \Delta}\| &\leq \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| \|g_\Delta\| \quad (\text{A.25}) \end{aligned}$$

Since $\|g_\Delta\| \leq \sqrt{k}$, we have

$$\|Q(\Delta)^{-1} g_{T \cup \Delta}\| \leq \sqrt{k} \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\|$$

Using (A.24) and (A.25), it gives

$$\begin{aligned} \|x - b_*\|_2 &\leq \gamma \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| + \|\lambda Q(\Delta)^{-1}(x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta})\| \\ &\quad + \|Q(\Delta)^{-1} A'_{T \cup \Delta} w\| \end{aligned}$$

A.3 Proof of Theorem 2

From the previous section we got the following conditions for the problem (2.3) to have a unique solution

$$\begin{aligned} ERC(T, \Delta) &\geq 0 \\ \gamma &\geq \frac{\|(y - Ac(\Delta))\|_2}{ERC(T, \Delta)} \end{aligned}$$

Under these conditions, the error, $\|x - \hat{x}\|_2$, of problem 2.3 will be

$$\begin{aligned} \|x - \hat{x}\|_2 &\leq \gamma \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| + \|\lambda Q(\Delta)^{-1} (x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta})\| \\ &\quad + \|Q(\Delta)^{-1} A'_{T \cup \Delta} w\| \end{aligned}$$

First, we obtain some conditions in terms of RIP and ROP constants that leads to $ERC(T, \Delta) \geq 0$.

$$ERC(T, \Delta) = (1 - \max_{i \notin T \cup \Delta} \|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1)$$

Let's bound the term $\|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1$. We know that for an arbitrary vector a we have $\|a\|_1 \leq \sqrt{|a|} \|a\|_2$, it gives

$$\|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1 \leq \sqrt{k} \|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_2 \leq \sqrt{k} \|(A'_\Delta M A_\Delta)^{-1}\|_2 \|A'_\Delta M A_i\|_2$$

We can derive that $\|A'_\Delta M A_i\|_2 \leq \theta_{k,1} + \frac{\theta_{u,k} \theta_{u,1}}{1 - \delta_u + \lambda}$ and if $\lambda \geq \delta_u + \frac{\theta_{u,k}^2}{1 - \delta_k} - 1$ then $\|(A'_\Delta M A_\Delta)^{-1}\|_2 \leq 1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}$. Applying these relations, it gives

$$\|(A'_\Delta M A_\Delta)^{-1} A'_\Delta M A_i\|_1 \leq \sqrt{k} \frac{\theta_{k,1} + \frac{\theta_{u,k} \theta_{u,1}}{1 - \delta_u + \lambda}}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}}$$

In other word

$$ERC(T, \Delta) \geq H_1(u, k, \lambda)$$

where

$$H_1(u, k, \lambda) = 1 - \sqrt{k} \frac{\theta_{k,1} + \frac{\theta_{u,k} \theta_{u,1}}{1 - \delta_u + \lambda}}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}} \quad (\text{A.26})$$

It means, for having $ERC(T, \Delta) \geq 0$ it is sufficient to have $H_1(|T|, |\Delta|, \lambda) \geq 0$. Now let's find the conditions that makes $H_1(|T|, |\Delta|, \lambda) \geq 0$.

$$1 - \sqrt{k} \frac{\theta_{k,1} + \frac{\theta_{u,k}\theta_{u,1}}{1-\delta_u+\lambda}}{1 - \delta_k - \frac{\theta_{u,k}^2}{(1-\delta_u+\lambda)}} \geq 0$$

Multiplying both sides with $1 - \delta_k - \frac{\theta_{u,k}^2}{(1-\delta_u+\lambda)}$ leads to

$$1 - \delta_k - \sqrt{k}\theta_{k,1} - \frac{\theta_{u,k}^2 + \sqrt{k}\theta_{u,k}\theta_{u,1}}{1 - \delta_u + \lambda} \geq 0$$

If $1 - \delta_k - \sqrt{k}\theta_{k,1} > 0$ we need

$$1 - \delta_u + \lambda \geq \frac{\theta_{u,k}^2 + \sqrt{k}\theta_{u,k}\theta_{u,1}}{1 - \delta_k - \sqrt{k}\theta_{k,1}}$$

$$\lambda \geq -1 + \delta_u + \frac{\theta_{u,k}^2 + \sqrt{k}\theta_{u,k}\theta_{u,1}}{1 - \delta_k - \sqrt{k}\theta_{k,1}}$$

Thus, $H_1(u, k, \lambda) > 0$ holds when

- $1 - \delta_k - \sqrt{k}\theta_{k,1}$
- $\lambda \geq g_2(u, k)$ where

$$g_2(u, k) = -1 + \delta_u + \frac{\theta_{u,k}^2 + \sqrt{k}\theta_{u,k}\theta_{u,1}}{1 - \delta_k - \sqrt{k}\theta_{k,1}} \quad (\text{A.27})$$

In the next step, our aim is to bring the inequality (A.23) in terms of RIP and ROP constants. Notice that

$$c(\Delta) = Q(\Delta)^{-1}(A'_{T \cup \Delta} y + \lambda \hat{\mu}_{T \cup \Delta}) \quad (\text{A.28})$$

Here we are going to find an upper bound for $\|(y - Ac(\Delta))\|_2$

$$\|(y - Ac(\Delta))\|_2 \leq \lambda \|A'_i A_{T \cup \Delta}\| \|Q(\Delta)^{-1}\| \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + \|(A'_i - A'_i A_{T \cup \Delta} Q(\Delta)^{-1} A'_{T \cup \Delta}) w\| \quad (\text{A.29})$$

In above inequality we have $\|Q(\Delta)^{-1}\|_2$. Since $\|Q(\Delta)^{-1}\|_2 \leq (\lambda_{\min}(Q(\Delta)))^{-1}$, we calculate the $\lambda_{\min}(Q(\Delta))$ in term of RIP constants.

$$\lambda_{\min}(Q(\Delta)) = \tag{A.30}$$

$$\begin{aligned} & \lambda_{\min} \left(\begin{bmatrix} A'_T A_T + \lambda I_T & A'_T A_\Delta \\ A'_\Delta A_T & A'_\Delta A_\Delta \end{bmatrix} \right) = \\ & \lambda_{\min} \left(\begin{bmatrix} A'_T A_T + \lambda I_T & 0 \\ 0 & A'_\Delta A_\Delta \end{bmatrix} \right) + \lambda_{\min} \left(\begin{bmatrix} 0 & A'_T A_\Delta \\ A'_\Delta A_T & 0 \end{bmatrix} \right) \geq \\ & \lambda_{\min} \left(\begin{bmatrix} A'_T A_T + \lambda I_T & 0 \\ 0 & A'_\Delta A_\Delta \end{bmatrix} \right) + \lambda_{\min} \left(\begin{bmatrix} 0 & A'_T A_\Delta \\ A'_\Delta A_T & 0 \end{bmatrix} \right) \end{aligned}$$

We bound the first and the last term in above inequality in the following corollaries.

Corollary 1. $\min(1 - \delta_u + \lambda, 1 - \delta_k) \leq \lambda_{\min} \left(\begin{bmatrix} A'_T A_T + \lambda I_T & 0 \\ 0 & A'_\Delta A_\Delta \end{bmatrix} \right)$

Corollary 2.

$$-\theta_{u,k} \leq \lambda_{\min} \left(\begin{bmatrix} 0 & A'_T A_\Delta \\ A'_\Delta A_T & 0 \end{bmatrix} \right)$$

Proof:

$$\lambda_{\min} \left(\begin{bmatrix} 0 & A'_T A_\Delta \\ A'_\Delta A_T & 0 \end{bmatrix} \right) = \min_{\|x_1\|^2 + \|x_2\|^2 = 1} [x_1 x_2] \begin{bmatrix} 0 & A'_T A_\Delta \\ A'_\Delta A_T & 0 \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} \tag{A.31}$$

$$\tag{A.32}$$

$$= \min_{\|x_1\|^2 + \|x_2\|^2 = 1} x_1 A'_T A_\Delta x'_2 + x_2 A'_\Delta A_T x'_1 = 2x_1 A'_T A_\Delta x'_2 \geq -2\theta_{u,k} \|x_1\| \|x_2\|$$

Since $\|x_1\|^2 + \|x_2\|^2 = 1$ then we can conclude that $\|x_1\| \|x_2\|$ is maximum when $\|x_1\| = \|x_2\| = \frac{1}{\sqrt{2}}$ so we have $\|x_1\| \|x_2\| = \frac{1}{2}$. By using corollary 2 and corollary 1 and the assumption that $\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k} > 0$, we obtain

$$\|Q(\Delta)^{-1}\| \leq \frac{1}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \tag{A.33}$$

Notice that for having $\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k} > 0$ we should at least have $1 - \delta_k - \theta_{u,k} > 0$.

Employing (A.33), we can rewrite (A.29) as follow

$$\begin{aligned} \|y - Ac(\Delta)\| &\leq \lambda \frac{\theta_{(u+k,1)}}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + \\ &\quad \left(1 + \frac{\theta_{u+k,1}(\sqrt{1 + \delta_{u+k}})}{(\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k})}\right) \rho \end{aligned} \quad (\text{A.34})$$

Using (A.34) and (A.26), we can rewrite (A.23) as

$$\gamma \geq H_2(u, k, \|x_{T \cup S} - \hat{\mu}_{T \cup \Delta}\|) \quad (\text{A.35})$$

where

$$H_2(u, k) = \frac{\lambda \frac{\theta_{(u+k,1)}}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \|x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta}\| + \left(1 + \frac{\theta_{u+k,1}(\sqrt{1 + \delta_{u+k}})}{(\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k})}\right) \epsilon}{1 - \sqrt{k} \frac{\theta_{k,1} + \frac{\theta_{u,k} \theta_{u,1}}{1 - \delta_u + \lambda}}{1 - \delta_k - \frac{\theta_{u,k}^2}{(1 - \delta_u + \lambda)}}} \quad (\text{A.36})$$

We have already found the conditions in terms of RIP and ROP constants which the unique solution for the problem (2.3) is obtained. We summarize the conditions which we have already obtained as follow

- $1 - \delta_k - \sqrt{k} \theta_{k,1}$
- $\lambda \geq g_2(u, k)$
- $\gamma \geq H_2(u, k, \|x_{T \cup S} - \hat{\mu}_{T \cup \Delta}\|)$ where g_2 and H_2 has been defined in and .

In the rest of this section we are going to get an explicit form for error bound. From the 2 we got the following error bound for the problem (2.3)

$$\begin{aligned} \|x - \hat{x}\|_2 &\leq \gamma \sqrt{|\Delta|} \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| + \|\lambda Q(\Delta)^{-1} (x_{T \cup \Delta} - \hat{\mu}_{T \cup \Delta})\| \\ &\quad + \|Q(\Delta)^{-1} A'_{T \cup \Delta} w\| \end{aligned} \quad (\text{A.37})$$

We have

$$\sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \leq \sqrt{\frac{\theta_{u,k}^2}{(1 - \delta_u + \lambda)^2} + 1}$$

and If $\lambda \geq \delta_u - 1 + \frac{\theta_{u,k}^2}{1 - \delta_k}$

$$\|(A'_\Delta M A_\Delta)^{-1}\| \leq \frac{1}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}}$$

So we can conclude

$$\sqrt{|\Delta|} \sqrt{\|(A'_T A_T + \lambda I_T)^{-1} A'_T A_\Delta\|^2 + 1} \|(A'_\Delta M A_\Delta)^{-1}\| \leq \sqrt{k} \sqrt{\frac{\theta_{u,k}^2}{(1 - \delta_u + \lambda)^2} + 1} \cdot \frac{1}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}} \quad (\text{A.38})$$

We saw that if $\lambda \geq \delta_u - 1 + \frac{\theta_{u,k}^2}{1 - \delta_k}$ and $1 - \delta_k - \theta_{u,k} \geq 0$ holds, then we can apply (A.33) and (A.38) to rewrite (A.37) as

$$\|x - \hat{x}\| \leq \gamma f_1(u, k, \lambda) + f_2(u, k, \lambda) \|x_{T \cup \Delta} - \hat{x}_{T \cup \Delta}\| + f_3(u, k, \lambda) \|w\|_2$$

where

$$f_1(u, k, \lambda) = \sqrt{k} \sqrt{\frac{\theta_{u,k}^2}{(1 - \delta_u + \lambda)^2} + 1} \cdot \frac{1}{1 - \delta_k - \frac{\theta_{u,k}^2}{1 - \delta_u + \lambda}} \quad (\text{A.39})$$

$$f_2(u, k, \lambda) = \frac{\lambda}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \quad (\text{A.40})$$

$$f_3(u, k, \lambda) = \frac{\sqrt{1 + \delta_{u+k}}}{\min(1 - \delta_u + \lambda, 1 - \delta_k) - \theta_{u,k}} \quad (\text{A.41})$$

A.4 Proof of Theorem 4

First, recall that $u_t := |T_t|$ and $k_t := |\Delta_t|$. The proof follows using induction. Using condition 5 of the theorem, the claim holds for $t = 0$. This proves the base case. For the induction step, assume that the claim holds at $t - 1$, i.e. $|T_{t-1}| \leq S_0, |\Delta_{t-1}| \leq S_1$ and $\|x_{t-1} - \hat{x}_{t-1}\| \leq C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma$. Using these assumptions we prove that the claim holds at t .

First, notice that condition 2 of theorem states that $\gamma = C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma + \sqrt{2dS_a r}$. We claim that under conditions 1 and 2 of theorem we have $\|x_t - \hat{x}_{t-1}\| \leq \gamma$. This is true since by Proposition 2 and the assumption of induction, we have $\|x_t - \hat{x}_{t-1}\| \leq \|x_{t-1} - \hat{x}_{t-1}\| + \sqrt{2dS_a r} \leq C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma + \sqrt{2dS_a r} = \gamma$. Note that $D_{S_0, S_1} < 1$ is necessary for having condition 2. It can be shown that $D_{S_0, S_1} < 1$ is equivalent to $1 - \delta_{2S_1} - \theta_{S_1, 2S_1} - 2(2 + \frac{(\sqrt{2}+1)\theta_{S_0, S_1}}{1 - \delta_{S_0}})\theta_{S_0, 2S_1} > 0$. This holds since condition 1 holds.

Now for the rest of proof if we show that $u_t \leq S_0$ and $k_t \leq S_1$, then Theorem 1 can be applied and we are done. That is because by condition 1, we can show that $\delta_{u_t} < 1$, $1 - \delta_{2k_t} - \theta_{k_t, 2k_t} > 0$ which are the first two requirements for applying Theorem 1. Condition

2 and the above discussion ensures that the third requirement of Theorem 1 holds.

First, notice that employing Proposition 1 and condition 3 and 4 of the theorem and by the assumption that $\|x_{t-1} - \hat{x}_{t-1}\| \leq C_{S_0, S_1} \epsilon + D_{S_0, S_1} \gamma$, we can conclude that at time $t-1$, all the elements greater or equal to $d_0 r$ will be get detected in the support update step, i.e. when computing \tilde{T}_{t-1} . Thus, the missed set, $|\tilde{\Delta}_{t-1}| \leq (2d_0 - 2)S_a$. Also, notice that by Proposition 2 and the assumption of induction we can conclude that at time $t-1$ no zero value element of x_{t-1} will be get detected as an element of \tilde{T}_{t-1} , in other word $\tilde{\Delta}_{e,t-1} = 0$. From Algorithm 1 we remember that $T_t = \tilde{T}_{t-1}$ and $u_t = |\tilde{T}_{t-1}|$. Since $\tilde{T}_{t-1} = N_{t-1} \cup \tilde{\Delta}_{e,t-1} \setminus \tilde{\Delta}_{t-1}$, we have $u_t = |\tilde{T}_{t-1}| \leq |N_{t-1}| + |\tilde{\Delta}_{e,t-1}|$. We know that $|N_{t-1}| = S_0$ and $|\Delta_{e,t}| = 0$. Hence, it implies that $u_t \leq S_0$.

Also, $\Delta_t = N_t \cap \tilde{T}_{t-1}^c = (N_{t-1} \cup A_t) \cap R_t^c \cap \tilde{T}_{t-1}^c \subseteq (\tilde{\Delta}_{t-1} \cup A_t) \cap R_t^c$. Here we have used the facts that $N_t = (N_{t-1} \cup A_t) \cap R_t^c$ and $\tilde{\Delta}_{t-1} = N_{t-1} \cap \tilde{T}_{t-1}^c$. So we have $\Delta_t \subseteq (\tilde{\Delta}_{t-1} \cup A_t) \cap R_t^c \subseteq (S_{t-1}(d_0) \cup A_t) \cap R_t^c$. Therefore, $k_t = |\Delta_t| \leq |S_{t-1}(d_0)| + |A_t| - |R_t| = (2d_0 - 2)S_a = S_1$.

BIBLIOGRAPHY

- [1] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Th.*, vol. 51(12), pp. 4203 – 4215, Dec. 2005.
- [2] D. Donoho, “Compressed sensing,” *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [4] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [5] E. Candes, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus de l’Academie des Sciences, Paris, Serie I*, pp. 589–592, 2008.
- [6] E. Candes and T. Tao, “The dantzig selector: statistical estimation when p is much larger than n ,” *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [7] D. Needell and J. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [8] D. Donoho, I. Drori, Y. Tsaig, and J. Starck, *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit*. Department of Statistics, Stanford University, 2006.

- [9] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [11] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM review*, pp. 129–159, 2001.
- [12] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, “Model-based compressive sensing,” *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [13] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [14] P. Schniter, L. Potter, and J. Ziniel, “Fast bayesian matching pursuit,” in *Information Theory and Applications Workshop, 2008*. IEEE, 2008, pp. 326–333.
- [15] D. Wipf and B. Rao, “Sparse bayesian learning for basis selection,” *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2153–2164, 2004.
- [16] N. Vaswani and W. L., “Modified-cs: Modifying compressive sensing for problems with partially known support,” *IEEE Transactions On Signal Processing*, September 2010.
- [17] F. Raisali and N. Vaswani, “Stability (over time) of regularized modified cs(noisy) for recursive causal sparse reconstruction,” in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, march 2011, pp. 1 –6.
- [18] N. Vaswani, “Kalman filtered compressed sensing,” in *ICIP*, 2008.
- [19] —, “Analyzing least squares and kalman filtered compressed sensing,” in *ICASSP*, 2009.
- [20] —, “Ls-cs-residual (ls-cs): compressive sensing on least squares residual,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4108–4120, 2010.

- [21] M. S. Asif and J. Romberg, “Dynamic updating for sparse time varying signals,” in *CISS*, 2009.
- [22] A. Carmi, P. Gurfil, and D. Kanevsky, “A simple method for sparse signal recovery from noisy observations using kalman filtering,” in *IBM Technical report*, December 2008.
- [23] D. Angelosante, G. Giannakis, and E. Grossi, “Compressed sensing of time-varying signals,” in *Digital Signal Processing, 2009 16th International Conference on.* IEEE, 2009, pp. 1–8.
- [24] R. von Borries, C. J. Miosso, and C. Potes, “Compressed sensing using prior information,” in *IEEE Intl. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2007.
- [25] A. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, “Weighted l1 minimization for sparse recovery with prior information,” in *ISIT*, 2009.
- [26] N. Vaswani, “Stability (over time) of modified-cs and ls-cs for recursive causal sparse reconstruction,” in *Allerton*, 2010.
- [27] J. A. Tropp, “Just relax: Convex programming methods for identifying sparse signals,” *IEEE Trans. Info. Th.*, pp. 1030–1051, March 2006.
- [28] J. Wright and Y. Ma, “Dense error correction via l1-minimization,” *IEEE Trans. Inf. Theor.*, vol. 56, pp. 3540–3560, July 2010. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2010.2048473>
- [29] W. Lu and N. Vaswani, “Modified bpdn for noisy compressive sensing with partially known support,” in *ICASSP*, 2010.
- [30] N. Vaswani and W. Lu, “Modified-cs: Modifying compressive sensing for problems with partially known support,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4595–4607, 2010.